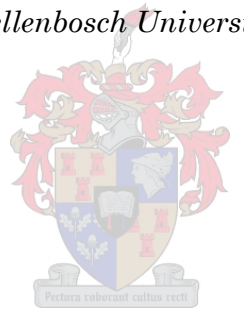
The utilisation of whole exome sequencing to dissect the genetic aetiology of familial Parkinson's disease in a South African Afrikaner family

By

Boiketlo Sebate

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (Human Genetics) in the Faculty of Medicine and Health Sciences at Stellenbosch University.



Supervisor: Associate Professor Soraya Bardien
Faculty of Medicine and Health Sciences
Department of Biomedical Sciences

Co-Supervisor: Dr Monique Williams
Faculty of Medicine and Health Sciences
Department of Biomedical Sciences

Co-Supervisor: Dr Ruben Cloete
South African National Bioinformatics Institute,
University of the Western Cape

April 2019

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification."

Signature :_____

Date : 30/11/2018

Abstract

Parkinson's disease (PD) is a complex neurodegenerative disorder, the aetiology of which is thought to be an interaction of genetic, biological and environmental factors. Its cardinal motor features, tremor, muscular rigidity, bradykinesia and abnormal gait occur relatively late in the course of disease, as a result of over 60% loss of dopaminergic neurons in the substantia nigra pars compacta. While most reported PD cases are sporadic, 5-10% of all cases are caused by several PD-causing genes including *Parkin, PINK1, LRRK2, SNCA, SYNJ1, DJ-1* and *EIF4G1*.

These PD genes were discovered using first-generation sequencing technologies which were expensive and time consuming. The development of high-throughput next generation sequencing technologies like whole exome sequencing (WES) has fast-tracked the discovery of disease-causing genetic mutations in various Mendelian disorders like PD. WES enables the screening of only the protein coding regions of the genome, to locate mutations which can interrupt cellular processes and lead to diseases. To date, WES has identified several PD-causing genes including *CHCHD2*, *VPS35* and *LRP10*, implicating the dysfunction of pathways regulating mitochondrial, lysosomal and synaptic function.

The aim of the present study is to combine WES technology and functional studies to identify a novel pathogenic mutation in a gene that could be implicated in the autosomal dominant form of PD in a South African Afrikaner family. To do this, a comprehensive filtration strategy was applied, combining various bioinformatic tools, in a step-wise analysis approach to assist in the filtration, interpretation and prioritisation of the NGS results. The bioinformatic tools used included SIFT, PolyPhen-2, MutationTaster, CADD, GERP++, Allen Brain Atlas, PANTHER and SWISS-MODEL. Once a single candidate gene was selected from the computational prioritisation, its protein expression was investigated in a disease relevant cell model using Western blotting.

WES was conducted on three affected individuals yielding over 20,000 variants each. Quality control (sequence alignment, alignment postprocessing, variant detection and quality evaluation) and filtering only the co-segregating non-synonymous variants through bioinformatics analysis yielded nine variants. Sanger sequencing was used to verify these variants, and four variants in the genes *ACTN3*, *CDC27, POU2F1* and *TUBB6* were found to be sequencing artefacts. Five variants in the genes *RFT1, NRXN2a, TEP1, CCNF* and *CFAP65* were found to be present in all four affected PD individuals of the family.

Only one variant, p.G849D in *NRXN2α,* fulfilled the various prioritisation criteria. The mutation was not present in the unaffected family members, in 671 South African PD patients and in 192 ethnically-matched controls. Multiple online population frequency databases also showed that the variant had not been previously reported in any population. Using a web-based database containing the exomes of 3000 patients with neurological disorders, 50 PD patients were identified with 24 other variants in the *NRXN2α* gene including indel mutations and premature stop codons. Amongst the five verified candidates, the p.G849D *NRXN2α* variant was predicted to be pathogenic across all four functional prediction tools with the highest Combined Annotation Dependent Depletion (CADD) score (29,50). It was also found to have a very high GERP++ score denoting that the level of evolutionary constraint acting on this site is predicted to be very high. The p.G849D *NRXN2α* amino acid change was the most severe, from a small, non-polar, side chain free amino acid, to Aspartic acid a larger, negatively charged amino acid. The homology modelling of the mutant vs wild-type revealed no change in the protein secondary structure but biochemically the substitution could lead to unwanted interactions with the neighbouring residues which could possibly affect the function and activity of the protein.

Also, NRXN2α was found to be highly expressed in the substantia nigra which is the main region of the brain affected in PD pathogenesis. It is associated with pathways related to calcium channel regulation, transmembrane signalling receptor activity, neuronal cell adhesion, synaptic organization and neuroligin family protein binding at the synapse, which makes it a plausible candidate gene for PD. Additionally, for functional studies, SH-SY5Y Neuroblastoma cells which are commonly used *in vitro* model for PD, were utilised. We investigated the endogenous NRXN2α levels in the SH-SY5Y cells and found that they produced detectable levels of the NRXN2α protein and were stably expressing this protein.

In summary, by integrating WES and *in vitro* studies, we identified p.G849D *NRXN2α*, a variant possibly associated with the autosomal dominant form of PD in a South African Afrikaner family. To our knowledge, this is the first report of an association between *NRXN2α* and PD. As a candidate, *NRXN2α* is well-suited for future functional mutant characterisation studies that will elucidate the impact of variants in this gene and their relative contribution to the disease phenotype. The further study of *NRXN2α* in PD may provide critical insight into novel disease mechanisms or genetic interactions with established PD mechanisms. Ultimately, this could potentially lead to development of improved therapeutic modalities for this debilitating disorder.

Opsomming

Parkinson se siekte (PS) is 'n komplekse neurodegeneratiewe siekte, waarvan die etiologie 'n interaksie is tussen genetiese, biologiese en omgewingsfaktore. Die kardinale motoriese simptome,word beskyf as spierstyfheid en bradykinesie. Hierdie abnormale gang word relatief laat in die verloop van die siekte waargeneem dit is as gevolg van die 60% verlies van dopaminerge neurone in die substantia nigra pars compacta. Terwyl die meeste PS-gevalle sporadies is word 5-10% van alle gevalle veroorsaak deur verskeie gene, insluitend *Parkin, PINK1, LRRK2, SNCA, SYNJ1, DJ-1* en *EIF4G1*.

Hierdie familiele PS-gene is ontdek deur gebruik te maak van eerste generasie sekwensie tegnologieë wat duur en tydrowend was. Die ontwikkeling van hoë-deursettingsvolgorde (NGS)-tegnologieë, soos die heel eksoom volgordebepaling (WES), het die ontdekking van siekte-veroorsakende genetiese mutasies in verskeie Mendeliese afwykings soos PS verspoedig. WES maak sifting van slegs die proteïenkoderende streke van die genoom om sodoende mutasies op te spoor wat sellulêre prosesse kan onderbreek en direk tot siektes kan lei. Tans het WES verskeie PS-veroorsakende gene geïdentifiseer, insluitend *CHCHD2, VPS35* en *LRP10* gene, wat geïmpliseer word in die disfunksie van paaie wat mitochondriale, lysosomale en sinaptiese funksie reguleer.

Die doel van die huidige studie is om WES tegnologie en funksionele studies te kombineer om 'n nuwe patogeniese mutasie in 'n geen te identifier wat in die outosomale dominante vorm van PS in 'n Suid-Afrikaanse Afrikaner-familie geïmpliseer kan word. Om dit te kan doen is 'n omvattende filtreertegnologie toegepas. Hierdie filtreer-tegnologieë kombineer doeltreffende nuwe bioinformatika-instrumente wat elkeen verskillende take aanspreek in 'n stap vir stap analise-benadering om te help met die filtrering, interpretasie en prioritering van die NGS-resultate. Die bioinformatiese gereedskap wat gebruik word sluit in SIFT, PolyPhen-2, MutationTaster, CADD, GERP ++, Allen Brain Atlas, PANTHER en SWISS-MODEL. Sodra 'n enkele kandidaat-geen uit die berekeningsprioritering gevind was, was ‘n funksionele analiese gedoen om die gene-uitdrukking in 'n siekte-relevante selmodel te ondersoek.

WES is uitgevoer op drie geaffekteerde individue wat meer as 20 000,00 variante lewer. Gehaltebeheer (volgordebelyning, belyningsprosessering, variantdeteksie en kwaliteitevaluering) en die filter van slegs die mede-segregerende, nie-sinoniem variante deur bioinformatika-analise het nege variante opgelewer. Sanger-volgordebepaling is gebruik om die nege WES-variante te bevestig, en vier variante in die gene *ACTN3, CDC27, POU2F1* en *TUBB6* was bevind as volgorde artefakte. Terwyl vyf variante in die

gene *RFT1, NRXN2α, TEP1, CCNF* en *CFAP65* gevind is, was dit teenwoordig in al vier die betrokke PS-individue van die familie.

Slegs een variant van die p.G849D-variant in *NRXN2α* het aan die verskillende prioritisasiekriteria voldoen. Die mutasie was nie teenwoordig in die onaangeraakte familielede in plaaslike pasiënte en kontroles nie. Dit is ook nie binne aanlyn bevolkingsfrekwensie databasisse gevind nie. Met behulp van 'n webgebaseerde databasis met die uitkomste van 3000 neurologiese pasiënte, is 50 PS pasiënte geïdentifiseer met 24 ander mutasies in die *NRXN2α*-geen, insluitend indel mutasies en voortydige stopkodons. Van die vyf vasgestelde kandidate word dit voorspel dat die p.G849D *NRXN2α*-variant patogenies is oor al vier funksionele voorspellingsgereedskap met die hoogste CADD-telling. Daar is ook gevind dat dit 'n baie hoë GERP++-telling het, wat aandui dat die vlak van evolusionêre beperking wat op hierdie webwerf optree, baie hoog is. Die p.G849D *NRXN2α*-aminosuurverandering was die ernstigste, 'n klein, nie-polêre, ketting vrye aminosuur, bewerkstellig ’n verandering tot Aspartiensuur wat 'n groter negatief gelaaide aminosuur is. Die homologie modellering van die mutant teenoor die wildtiepe het geen verandering in die proteïen sekondêre struktuur geopenbaar nie, maar biochemies kan die verandering lei tot ongewenste interaksies met die naburige residue wat moontlik die funksie en aktiwiteit van die proteïen kan beïnvloed.

Ook, NRXN2α was hoogs uitgedruk in Pars Reticulata van die substantia nigra, die hoofstreek van die brein wat in PS patogenese geraak word. Dit word geassosieer met paaie wat verband hou met kalsiumkanaalregulering, transmembraansein, reseptoraktiwiteit, neuronale adhesie, sinaptiese organisasie en neuroligin-familie-proteïenbinding. Vir funksionele studies is SH-SY5Y-selle wat 'n gevestigde en algemeen gebruikte *in-vitro*-model is vir PS , gebruik. Ons het die endogene NRXN2α-vlakke in die SH-SY5Y-selle ondersoek en gevind dat hulle waarneembare vlakke van die NRXN2α-proteïen produseer en hierdie proteïen stabiel uitdruk.

Deur integere-sekwensvolgorde en funksionele studies te integreer, het ons p.G849D *NRXN2α* geïdentifiseer in 'n variant wat moontlik met ‘ n outosomale dominante vorm van PS in 'n Suid-Afrikaanse Afrikaner-familie geassosieer kan word. Ons weet dit is die eerste verslag van 'n assosiasie tussen *NRXN2α* en PS. As ’n kandidaat is *NRXN2α* geskik vir toekomstige funksionele mutasie karakteriserings-studies wat die impak van variante in hierdie geen en hul relatiewe bydrae tot die siekte fenotipe sal verhelder. Die verdere studie van *NRXN2α* in PS kan kritiese insig gee in nuwe siekte meganismes of genetiese

interaksies met gevestigde PS meganismes. Dit kan moontlik lei tot die ontwikkeling van verbeterde terapeutiese modaliteite vir hierdie afwykende siekte.

Acknowledgements

"What we do for ourselves dies with us.

What we do for others and the world remains and is immortal..."

~ Sir Albert Bigelow Paine, 1901

This thesis is a culmination of the contribution of so many people. Their thoughts, energy, time, financial contribution, unplanned chats, sleepless nights, tears and prayers. To my supervisors (Prof Bardien, Dr Williams and Dr Cloete) who patiently and painstakingly walked me through this journey, I am because you were, I thank you. To my dear mother who taught me taught me that only perseverance will be rewarded, thank you so much. To the amazing lab mates who kept my spirits high with humour and sometimes a simple smile or hug, I appreciate you all. To the admin staff of the division whose contribution sometimes goes unnoticed, I saw you and thank you for all that you did for me. To the patients and their family who made this study possible, thank you so very much. Lastly, to my poor husband who listened to countless hours of "science talk" and now claims to have enough knowledge on genetics to write a review, I thank you too.

It's almost an injustice to thank but just a few, I therefore extend my heartfelt gratitude to everyone who in some way or form was a part of this journey. May the work done here not die with me but be a stepping stone on the search for answers and relief from this disease. The financial assistance of the National Research Foundation towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

"And whatsoever ye do in word or deed, do all in the name of the Lord Jesus, giving thanks to God"

List of Abbreviations

AAO	Age at onset
AD	Autosomal dominant
AP	Atypical Parkinsonism
AR	Autosomal recessive
BAM	Binary Alignment/Map
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CADD	Combined Annotation Dependent Depletion
CAF	Central Analytical Facility
cDNA	Complementary Deoxyribonucleic Acid
CLB	Cell Lysis Buffer
CNV	Copy number variation
DBS	Deep brain stimulation
DNA	Deoxyribonucleic Acid
EGF	Epidermal growth factor
ExAC	Exome Aggregation Consortium
gDNA	Genomic Deoxyribonucleic Acid
GMQE	Global Model Quality Estimation
HRM	High resolution DNA melting
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Lewy bodies
MAF	Minor allele frequency
MLPA	Multiplex Ligation-dependent Probe Amplification
MRI	Magnetic resonance imaging
MSA	Multiple System Atrophy
NGS	Next generation sequencing
NHGRI	National Human Genome Research Institute
NMS	Non-motor symptoms
NTC	No template control
NR	Not Reported
NTC	No template control
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDB	Protein Data Bank

PS	Parkinson se siekte
PSP	Progressive Supranuclear Palsy
QMEAN	Qualitative Model Energy Analysis
REM	Rapid Eye Movement
RNA	Ribonucleic Acid
RT	Reverse transcription
SA	South Africa
SAM	Sequence alignment/map
SDS	Sodium dodecyl sulphate
SIFT	Sorts Intolerant From Tolerant
SNP	Single nucleotide polymorphisms
SSA	Sub-Saharan Africa
TA	Annealing Temperature
TM	Melting Temperature
UPDRS	Unified Parkinson's Disease Rating Scale
UTR	Untranslated Region
WB	Western blot
WES	Whole exome sequencing
WGS	Whole genome sequencing

Table of Contents

List of Figures

List of Tables

Outline of Thesis

Chapter 1 provides background information on PD with emphasis on the advances in sequencing technologies and the PD genetics research that has been conducted in South Africa to date. The present study describes the scientific value of the work and it is followed by the main objectives of the project.

Chapter 2 is the Materials and Methods which describes how the experiments were conducted. It is written in two sections; the first section is of the investigation of the whole exome sequencing data using bioinformatics tools to estimate the impact of the detected single nucleotide variants. This is followed by filtration based on set criteria and the resulting process of candidate gene prioritisation. Sanger sequencing was used to validate the presence of WES variants discovered in the PD patients. Then last to be described is the determination of the variant frequencies in multiple control populations.

The second section was of the Methods and Materials is the investigation of the prioritised gene. This included determining the variant frequency in the local PD population as well as an international cohort of PD patients. The structural homology modelling of the wild-type and mutant protein were done, to determine the impact of the variant on protein folding and amino acid interactions. Multiple attempts of generating cloning inserts ensued, before functionally determination the endogenous protein levels of the prioritised gene in a neuroblastoma cell line.

Chapter 3 is the results section, it contains the findings of the study in the same sequential order as the methodologies. Documenting the WES analysis results, filtration criteria and the candidate prioritisation. And the following section having the results of all that was done to characterise the prioritised gene.

To conclude is Chapter 4, which has a detailed discussion on the findings of the study, the potential implications of these results, study limitations, and recommendations regarding further research.

Chapter One
1. Introduction and Literature Review

1.1. Introduction to Parkinson's Disease

1.1.1. Background

The clinical syndrome described as Parkinson's disease (PD) [OMIM #168600] was first documented in western medical literature by a London-based doctor, James Parkinson, in 1817. Dr. Parkinson wrote a medical case report entitled "An Essay on the Shaking Palsy" (Parkinson, 1817), and in his report he often referred to older findings of this "shaking palsy" by scholars before his time including the Italian artist, engineer and scientist Leonardo da Vinci (1489-1506) and the late Roman physician Claudius Galenus (1200 - 1350 Anno Domini) (Forno, 1996).

Preceding this account, PD as a disorder had been identified and treated in different parts of the world since ancient times. One clear description of this can be found as early as 300 BC in the ancient Indian system of medicine of Ayurveda, where a disorder named Kampavata is described that is consistent with parkinsonism in all of its components (Ovallath and Deepa, 2013). Since then PD has become the second most common neurodegenerative disorder, affecting 1-2% of the global population above the age of 65 years (de Rijk et al., 2000). Research has uncovered much about PD, but the underlying disease mechanisms remains a mystery.

1.1.2. Neuropathology of Parkinson's Disease

The major neuropathological symptoms of PD include, but are not restricted to, dopaminergic neuronal degeneration, demyelination, gliosis and depigmentation of the substantia nigra (Jankovic, 2008; Hauser et al., 2016). The substantia nigra is responsible for muscle tone, ease of movement and learned muscle patterns (LaRocco, 2015) and its depigmentation is due to the loss of neuromelanin in the dopaminergic neurons of the locus coeruleus (**Figure 1.1**). It is estimated that when the motor symptoms of PD emerge, approximately 50-70% of dopaminergic neurons have already been lost from the substantia nigra (Cheng et al., 2010).

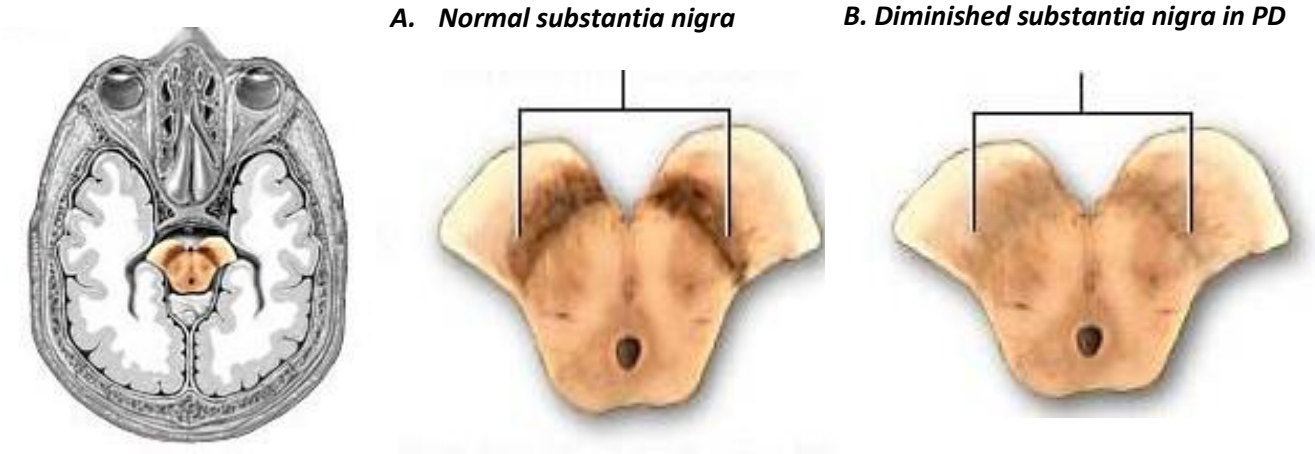


Figure 1.1 Transverse section of midbrain showing the difference in the substantia nigra of a healthy control and a PD patient. Shown on the far-left is the location of the substantia nigra in a transverse cross section of the brain **A**. Visibly dark normally distributed dopaminergic neurons in the substantia nigra of a healthy control. **B.** Severe depigmentation of the substantia nigra due to loss of neuromelaninated dopaminergic neurons of a PD patient. (Adapted with permission from A.D.A.M. https://medlineplus.gov/ency/imagepages/19515.htm)

Another major neuropathological symptom of PD is the presence of intracellular proteinaceous inclusions called Lewy bodies (LBs) and accumulations of alpha-synuclein filaments with granular material referred to as Lewy neurites (**Figure 1.2**) (Spillantini et al., 1998). Both these inclusions are found in brain regions such as the brainstem, cortical areas and locus ceruleus (Wakabayashi et al., 2007; Belin and Westerlund, 2008). Lewy pathology occurs in about 8–17% of neurologically normal people above the age of 60 (Gibb and Lees, 1988; Frigerio et al., 2011), this finding is termed incidental LBs. However, in PD patients the prevalence of these incidental LBs has been shown to increase aggressively with age (Gibb and Lees, 1988). It has been hypothesised that LB aggregation represents the pre-symptomatic phase of PD (Lacono et al., 2015).

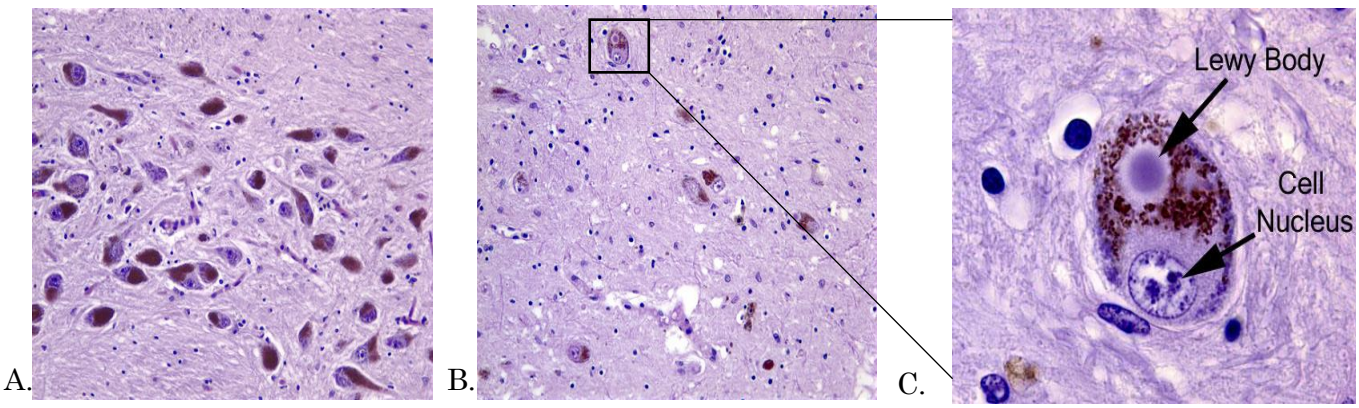


Figure 1.2 Histological section of the substantia nigra zona compacta of a healthy control and a PD patient. A. Immunohistochemical hematoxylin and eosin staining of zona compacta in a healthy control. **B.** Staining of the same region in a PD

patient shows the loss of pigmented neurons and neurons with the eosinophilic spherical LB. **C**. A magnified view of the pigmented neuron in the PD patient, containing eosinophilic cytoplasmic inclusions with a halo (Adapted with permission from http://neuropathology-web.org/chapter9/chapter9dPD.html).

Braak and colleagues investigated the neuropathology of PD in 41 patients (19 females, 22 males, aged 75,7 ±7,2 years) who had all been clinically diagnosed with PD and all had accumulations of alpha-synuclein filaments (they exhibited nigral LBs and severe loss of nigral dopaminergic neurons). From this study they proposed that, within the central nervous system PD begins as a synuclein pathology in non-dopaminergic structures of the lower brainstem or in the olfactory bulb (Braak et al., 2002). They noted and demonstrated that the varying degrees of synuclein pathology depicted a sequence of events that progressed to affect the substantia nigra and ultimately caused parkinsonism (Braak et al., 2003). This concept was further investigated and developed into a staging of the pathology of PD and has become widely accepted with its observations now required for a definitive diagnosis of PD (Braak et al., 2007) (Moore et al., 2005).

The Braak staging (**Figure 1.3**) proposes that in the earliest stage of PD, stage one, there is abnormal α-synuclein accumulation in either the medulla oblongata or the olfactory bulb. In stage two the LB pathology develops in the locus coeruleus in the pontine tegmentum of the brainstem. Stages three and four involves the development and increase of LBs as well as the severe loss of dopaminergic neurons in the substantia nigra/amygdala. It is at this stage of the pathology that the clinical signs of PD are visible, and it is then possible to diagnose PD by current clinical criteria. In stages five and six the synuclein pathology invades the neocortex and spreads into other the structures of the brain. Cell death occurs in the substantia nigra, both the motor and sensory areas of the brain are affected and the disease is at its most severe (Braak et al., 2002; Burke et al., 2008; Halliday et al., 2011).

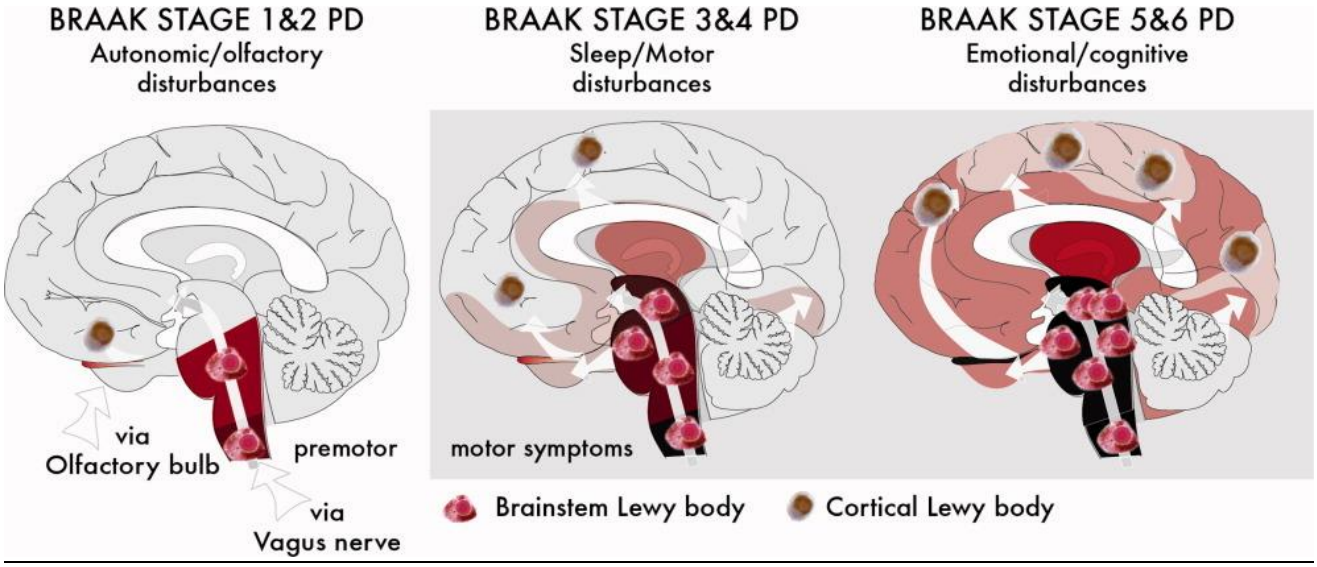


Figure 1.3 Stages in the evolution of PD-related pathology. Stylized representation of the Braak staging for Parkinson's disease showing the initiation sites in the medulla oblongata and olfactory bulb through to the later infiltration of Lewy pathology into the cortical regions (Halliday et al., 2011). Used with permission from John Wiley and Sons, Inc.

1.1.3. Motor and Non-motor Features of Parkinson's Disease

The motor features of PD typically begin subtly, becoming prominent over time with tremor being the initial symptom that is most widely recognised (Hauser et al., 2016). In his clinical report Dr. Parkinson described the motor symptoms that define PD as:

"Involuntary tremulous motion, with lessened muscular power in parts not in action, and even supported, with a propensity to bend the trunk forward, and to pass from a walking to a running pace, the senses and intellect being uninjured."(Parkinson, 1817).

Sixty years after the publication of this report, a French neurologist, Dr. Jean Martin Charcot, recognised the importance of Parkinson's observations and justly named the disease after Dr. Parkinson. Charcot further refined and expanded this early description of the disease by adding rigidity to what is now recognised as the four cardinal motor symptoms of PD; bradykinesia (slowness of movement), tremor (unintentional shaking), rigidity (muscular stiffness) and postural instability (balance impairment) (Kapur et al., 2012). Over time more motor features have been identified though classified as minor because they only occur in a small fraction of all PD patients. These include reduction in dexterity (fine motor skills), coordination, and decreased arm swing on the first-involved side (Antonini et al., 2015).

Dr. Parkinson also noted an array of symptoms that did not involve movement, coordination and physical tasks. These were later coined as non-motor symptoms (NMS)

and have become a hallmark of disease manifestation with some NMS preceding the motor symptoms by up to ten years. The most common being olfactory impairment, excessive production of saliva, sleep disturbances, dysphagia and autonomic dysfunction (constipation, bladder dysfunction and orthostatic hypotension) (Hauser et al., 2016).

In advanced PD, patients experience NMS related to mild cognitive impairment which range from dementia, psychosis, bradyphrenia (slowing in cognitive abilities), hallucinations and depression (Meireles and Massano, 2012). The recognition of NMS and their impact on the patient's quality of life has led to a significant shift in diagnostic and therapeutic approaches to the disease.

1.1.4. Differentiating Parkinson's Disease from Atypical Parkinsonian Syndromes.

Atypical Parkinsonism (AP) is a term describing a spectrum of neurodegenerative disorders that have some of the classic features of PD but have additional symptoms that are not typically present in PD and have a different pathology. AP disorders include Multiple System Atrophy (MSA), Progressive Supranuclear Palsy (PSP), and Corticobasal Degeneration (CBD) (Behnke et al., 2005). It is vital to differentiate PD from AP as the treatment, symptoms and prognosis for each disease is different. The phenotypic symptoms present in AP but absent in PD are the onset of motor symptoms occurring on both sides of the body, experiencing falls early in disease manifestation, dementia accompanied by early cognitive decline and significant visual complications (Colosimo et al., 2010).

The clinical differences between the two are that PD motor symptoms are typically alleviated by dopamine replacement in the form of Levodopa (LD) while AP patients exhibit poor or no response to LD (Hughes et al., 1992). This is because in PD the dopamine producing nerve cells in the substantia nigra are lost, however the nerve cells in the basal ganglia with dopamine receptors are preserved and capable of responding to treatment. Also, 5% to 10% of AP patients have lesions in other systems and a worse prognosis from disease onset (Trenkwalder et al., 2017).

1.1.5. Diagnosis and Treatment of Parkinson's Disease

Currently, the commonly accepted method of PD diagnosis is the application of the stringent United Kingdom PD Society Brain Bank criteria by a trained neurologist (Hughes et al., 1992). This criteria states that the patient must have bradykinesia and at least one of the following; muscular rigidity, rest tremor (4–6 Hz) or postural instability

(Alfred Goldscheider, 1898; Clarke et al., 2016). This diagnostic screening criteria can be accompanied by an array of ancillary tests, such as genetic testing, olfactory testing, magnetic resonance imaging (MRI), and dopamine-transporter single-photon-emission computed-tomography imaging, to solidify diagnostic accuracy (Tolosa et al., 2006).

There is currently no cure for PD nor is there any treatment that has definitively been shown to slow the progression of the disease. As a result, the treatments that have been developed are predominantly used to treat the various symptoms (Antonini et al., 2015). Most of the available medications are designed to help replace the dopamine deficiency that occurs in PD (Booker, 2012).

The gold standard or primary medication dispensed is LD. It acts directly on the brain replenishing depleted dopamine stores as it is a precursor to dopamine with the unique ability of crossing the blood-brain barrier. Although it is the most effective medication for symptomatic treatment of the motor symptoms, this motor symptom suppressive effect will eventually wear off with chronic LD administration and disease progression (LeWitt et al., 2016).

High frequency deep brain stimulation (DBS) is a surgical procedure for PD treatment that does not cure PD, but has been proven to diminish symptoms (Benabid et al., 2009). It involves implanting a small metal electrode into either the subthalamic nucleus, globus pallidus interna, or ventral lateral nucleus of the thalamus (Weaver, 2009). The advantages of DBS are that it improves symptoms like tremor, rigidity, and bradykinesia significantly, resulting in the decrease of prescribed LD doses by an average of 60% (Limousin et al., 1998; Groiss et al., 2009). The procedure is reversible and does not cause damage to brain tissue.

A disadvantage of the DBS surgery is the risk of developing a postoperative stroke caused by bleeding in the brain. This can result in permanent brain damage or death, though it is reported that fewer than one percent of PD patients who have undergone DBS surgery have experienced a stroke (Stefani et al., 2007; Zrinzo et al., 2011). Other complications of DBS surgery include infection, malfunction of the neurostimulator and movement of the electrode (Limousin et al., 1998; Follett et al., 2010).

1.1.6. Prevalence and Incidence of Parkinson's Disease

Prevalence and incidence are epidemiological measures of disease frequency in populations. The prevalence is a ratio of the number of existing cases of a disease in a population of a defined size (de Lau and Breteler, 2006). In contrast, incidence is a rate of the number of new reported cases of disease in a population of a defined size, during specific time period (de Lau and Breteler, 2006).

Globally, the prevalence of PD is predicted to be within the range of 41 per 100,000 people in the fourth decade of life (Pringsheim et al., 2014). In individuals aged 80 years and older it is predicted to be more than 1,900 people per 100,000 (Pringsheim et al., 2014; Cacabelos, 2017). Approximately 60,000 people were diagnosed with PD in the United States of America in 2002, and eight years later this number increased ten times to 630,000 (Obeso et al., 2000, Kowal et al., 2013). In 2000, Europe had a prevalence of 1,803 per 100,000 people above the age of 65 years, and this increased significantly with age to 2599 per 100,000 people between 85 to 89 years (de Rijk et al., 2000). A review published by Dorsey et al., (2007) estimated the number of individuals suffering from PD in ten of the most populated countries in the world. This was reported to be between 4,1 and 4,6 million people in 2005 and they also projected that the global PD prevalence is expected to increase twofold by the year 2030 (Dorsey et al., 2007).

Compared to other regions, there is a lack of epidemiological studies in Sub-Saharan Africa (SSA) and the studies that have been conducted are either very old or have a very small sample size. The prevalence of PD in the countries of SSA is perceived to be considerably lower than that reported in the more developed countries. Previously Africa as a continent had been estimated to have similar rates of prevalence as those of China and Japan ranging between 88 to 133 per 100,000 people (Zhang and Román, 1993). A review of SSA prevalence studies conducted in 2013 estimated an even lower prevalence of PD in SSA, between 7 to 20 per 100,000 people (Blanckenberg et al., 2013). To date, there have only been 12 prevalence studies conducted in SSA (Cosnett, 1973; Lombard and Gelfand, 1978; Osuntokun et al., 1987; Schoenberg et al., 1988; Haimanot et al., 1990; Benamer et al., 2008; Dotchin et al., 2008; Winkler et al., 2010; Akinyemi, 2012; Femi et al., 2012; Cubo et al., 2014; Kisoli et al., 2015) which is composed of 51 countries with a population of a billion citizens (Department of Economic and Social Affairs, 2016).

In SA, it has been 45 years since the first and only prevalence study was conducted in a neurology unit of a Durban public hospital cohort in 1973. The calculated prevalence of PD in the black SA population was 150 per 100,000 people, in the SA Indian population it

was 1,260 per 100,000 people, while the prevalence of PD in the Caucasian SA population was 2,310 per 100,000 people (Cosnett, 1973).

The annual incidence rate of PD globally differs greatly between populations ranging from 5 per 100,000 to 346 per 100,000 people per year in European countries (Campenhausen et al., 2005) versus in China where it is between 1,5 per 100,000 to 8,7 per 100,000 people per year (Zou et al., 2015).The only reported incidence rate estimate for PD in a SSA population came from a study done in North Africa that had included the Arabs of Eritrea as part of the North African patient cohort (Benamer et al., 2008). The calculated annual incidence rates of PD in the North African Arabs between 1982 and 1984 was reported to range from 4,5 per 100,000 people, to 20 per 100,000 people per year (Benamer et al., 2008). Multiple studies have inferred that PD might be more common in the Northern African Arabic population than the rest of SSA, and therefore this may not reflect the overall incidence. More data is required to confirm this (Okubadejo, 2008; Bardien et al., 2010; Blanckenberg et al., 2013).

1.2. Aetiology and Risk Factors for Parkinson's Disease

1.2.1. Sporadic Versus Familial Parkinson's Disease

PD can present in two forms, namely sporadic PD and familial PD. Familial PD is less prevalent as approximately only 10% of patients report a positive family history (Klein and Westenberger, 2012), but its causation is better understood when compared to sporadic PD. The discovery during the last two decades of gene mutations that are directly linked to acquiring heritable forms of PD [autosomal dominant (AD), autosomal recessive (AR), and sex chromosome linked] has confirmed the role of genetics in the development of PD. These mutations have also provided vital clues in understanding the molecular pathogenesis of PD (Klein and Westenberger, 2012).

Sporadic PD is thought to be an interaction between genetic factors, environmental toxins, and oxidative stress which together ultimately impact neuronal loss (Lesage et al., 2006). The mechanisms believed to be involved in the development of sporadic PD include the aggregation of abnormal proteins (Langston, 2006), damage to cellular membranes, mitochondrial dysfunction (Winklhofer and Haass, 2010), impaired transportation in neural synaptic clefts (Belin and Westerlund, 2008), and ultimately programmed cell death (Perier et al., 2011).

1.2.2. Environmental and Demographic Factors of Parkinson's Disease

In the absence of a clear genetic cause for the majority of cases, environmental influences are thought to play an important role in the development of PD (Di Monte et al., 2002). These environmental risk factors include pesticides, organic solvents, metals, and air pollutants (Dick et al., 2007). There have also been environmental influences that have been found to exert beneficial effects against PD development, these include polyphenols, curcumin, green tea, coffee and nicotine (Abdel-Salam, 2019).

1.2.2.1. Toxins

A study by Langston and colleagues reported parkinsonism in a group of drug addicts following injection of self-produced drugs containing N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (Langston et al., 1983). This toxin has been found to selectively damage cells in the substantia nigra by inhibiting Complex I (NADPH-ubiquinone oxidoreductase I) of the electron transport chain, which leads to impaired ATP production, elevated intracellular calcium concentrations an increase in reactive oxygen species, and ultimately results in cell death (Watanabe et al., 2005).

A well-researched form of toxin exposure is farming or rural living, which are both associated with a greater risk of PD (Seidler et al., 1996; Liou et al., 1997). Two toxins that has become evident through farming practices are the herbicide paraquat and the pesticide rotenone. Paraquat causes oxidative stress and rotenone inhibits mitochondrial complex I (Langston et al., 1983; Sherer et al., 2007). In multiple experimental models of PD, both toxins have been shown to induce the loss of dopaminergic neurons in the substantia nigra(Cacabelos, 2017). Over the years, other neurotoxins have emerged from the agricultural occupation including organophosphates, organochlorine, metals and carbamate (Hertzman et al., 1990; Semchuk et al., 1991, 1992, 1993; Di Monte et al., 2002; Dick et al., 2007).

1.2.2.2. Caffeine

The effect of caffeine on PD risk is well studied and it has been found that the consumption of caffeine rich beverages like green tea amongst Chinese or coffee amongst Caucasians results in a protective effect against PD (Ross et al., 2000). This finding is consistent with the results of studies conducted in Germany and Sweden (Ascherio, et al., 2001). Caffeine is a known antagonist to adenosine, which is a signalling molecule that binds to receptors specifically found on dopaminergic neurons of the basil ganglia to reduce the release of

dopamine. By inhibiting adenosine binding, caffeine temporarily induces the release of dopamine (Muñoz and Fujioka, 2018). This is one mechanism caffeine may employ to prevent neurodegeneration damage and decrease bradykinesia and rigidity in PD patients (Postuma et al., 2012).

1.2.2.3. Cigarettes

Despite the adverse health effects of cigarette smoking, a large number of studies have consistently found that smokers have a lower incidence of PD than non-smokers (Baumann et al., 1980; Grandinetti et al., 1994; Di Monte et al., 2002). This implies that smoking could be a protective factor for PD. It was found that the more years of smoking, rather than increased number of cigarettes smoked per day, was linked closely to lowering the risk of PD amongst smokers (Gorell et al., 1999; Hellenbrand et al., 1997).

It is not clear why this occurs, but it has been found experimentally that in animal models nicotine interacts with receptors to protect dopaminergic neurons (Quik and Kulak, 2002). A number of studies has even shown the protective effects of nicotine against toxin, MPTP, further strengthening the nicotine neuroprotective hypothesis (Group, 2006, Muñoz and Fujioka, 2018).

There have been other findings regarding cigarette smoking and PD. In case-control studies researchers observed a significantly lower risk of PD for smoking as early as 15 to 20 years before symptom onset, but not for smoking 25 years before onset (Hernán et al., 2001, Thacker et al., 2007). This suggests a temporal relationship between cigarette smoking and PD risk, with a lower risk amongst current smokers, than in former smokers.

1.2.2.4. Age

PD occurrence has been shown to directly increase with advancing age, making age the strongest risk factor for disease development (Driver et al., 2009; Reeve et al., 2014; Cacabelos, 2017). The mean age at onset of PD is approximately 60 years, with only 4% of all patients developing the early-onset form of PD before the age of 50 (Bower et al., 1999). It has been established that with advancing age there is naturally a decline in dopamine neuron function and an accumulation of dysfunctional cell organelles of neurons (Müller and Pedigo, 1994; Thal et al., 2004). However, in PD it is thought that the normal course of ageing is disrupted to such an extent that it results in a significant loss of dopamine producing neurons and thereafter PD symptoms (Calne and Langston, 1983). As populations grow older than before, due to advances in medicine and the abundance of

food, age-related illnesses like PD will become more common (Campbell and Tishkoff, 2008; Ramsay et al., 2011).

1.2.2.5. Gender and Hormones

Studies have reported a higher occurrence of PD in men than in women. The overall annual incidence rate of PD in females 40 years of age and older is 37,55/100,000 compared to 61,21/100,000 people in males within the same age range (Hirsch et al., 2016). One study has suggested that observable higher risk of PD in males than in females is a result of the hormone oestrogen. It is highly expressed in females and thought to exert neuroprotective benefits (Saunders-Pullman, 2003).

A study by Haaxma et al., (2007) conducted on 253 PD patients found several gender differences between patients. The female patients tended to be older than males at the age at onset. They more often presented a tremor dominant form of PD which has been associated with a slower disease progression. Also at onset, female PD patients had higher levels of striatal dopamine binding than men suggesting a more benign PD phenotype in females when compared to males (Haaxma et al., 2007).

1.2.3. Genetic Risk Factors of Parkinson's Disease

Research on the genetic causes of PD over the past two decades has led to the discovery of several PD-causing or associated genes. Summarised in **Table 1.1** are the established PD genes with mutations proven to be pathogenic.

The AD PD genes identified include *SNCA* (PARK1,4), *LRRK2* (PARK8), *VPS35* (PARK17), *EIF4G1* (PARK18) and *LRP10*. The characteristics shared by these AD PD genes are typical PD, a good response to LD in the early stages of disease progression, LB formation and majority of these genes result in late-onset PD. The differences in associated disease phenotypes includes dementia, cognitive decline and sleep abnormalities in PD patients with *SNCA* mutations (Zarranz et al., 2004; Zarranz et al., 2005), while patients with *LRRK2* mutations show amyotrophy and/or severe dystonia (Zimprich et al., 2004; Gilks et al., 2005). *VPS35* patients tend to have tremor dominant parkinsonism accompanied by muscle spasms and depression (Wider et al., 2008; Vilariño-Güell et al., 2011),while PD patients with *EIF4G1* mutations tend to exhibit mild disease with a few having dementia (Chartier-Harlin et al., 2011; Schulte et al., 2012). Patients who have mutations in the recently discovered *LRP10* gene exhibited typical PD with dementia and in a small number of cases there were unique features observed such as

supranuclear gaze palsy and ideomotor apraxia (an inability to perform a gesture using a limb upon verbal command or by imitation) (Quadri et al., 2018).

The AR PD genes include *DJ-1* (PARK7), *FBXO7* (PARK15), *Parkin* (PARK2), *PINK1* (PARK6) and *SYNJ1 (PARK20)* (**Table 1.1**). The characteristics shared by these AR PD genes are typical PD, majorly juvenile to early disease onset and LB pathology is generally absent. The differences in phenotype between these genes are that *PARK2* and *PINK1* patients exhibit very similar disease phenotype with lower limb dystonia and no cognitive NMS (Ahlskog, 2009; Samaranch et al., 2010), whereas *DJ-1* patients, although similar in phenotype to *PARK2* and *PINK1* patients, also exhibit early dementia and amyotrophy (Lesage and Brice, 2009). Patients with *SYNJ1* mutations tend to have dystonia with rapid cognitive decline and seizures (Krebs et al., 2013). *FBXO7* patients exhibit typical PD with pyramidal signs such as hyperreflexia, limb atrophy and spasticity (Di Fonzo et al., 2009).

Table 1.1 List of genes conclusively implicated in Mendelian forms of PD.

Gene (locus)	Typical symptoms	Mutations	Average age at onset (Years)	Reference
Autosomal dominant				
***EIF4G1* (PARK18)**	Late-onset Lewy bodies present	p.R1205H	≥50	Chartier-Harlin et al. 2011 Nichols - 2015 Tucci et al 2012
***LRRK2* (PARK8)**	Late-onset Good response to Levodopa Lewy bodies present Slow progression	p.G2019S p.R1441G p.R1441C p.R1441H p.I2020T p.Y1699C p.R1437H	≥50	Zimprich et al. 2004 Ozelius et al. 2006 Ujiie et al. 2012
***LRP10* (No PARK locus)**	Late-onset Lewy bodies present Progression	p.G603R p.P699S p.Y307N p.R533L p.R235C p.1424+5delG (intronic) p.N517del n.A212Sfs17	≥60	(Quadri et al., 2018a)
***SNCA* (PARK1,4)**	Variable onset Good response to Levodopa Dementia Cognitive decline	p.A53T p.A30P p.E46K p.H50Q p.G51D	Variable onset	Polymeropoulos et al. 1996 Klein et al.2009

Gene (locus)	Typical symptoms	Mutations	Average age at onset (Years)	Reference
	Lewy bodies present Fast progression	p.A53E		
***VPS35* (PARK17)**	Late-onset Good response to Levodopa Slow progression	p.D620N p.P316S	≥50	Vilariño-Güell et al. 2011
Autosomal recessive				
***DJ-1* (PARK7)**	Early-onset Lewy bodies absent Slow progression Psychiatric features	p.A39S p.E64D p.D149A p.Q163L p.L166P p.M261I dup168-185	≥35	Bonifati et al 2003
***FBXO7* (PARK15)**	Juvenile or early-onset Pyramidal signs Fast progression	p.T22M p.R378G p.R498X	≤20	Di Fonzo et al. 2009
***Parkin* (PARK2)**	Early-onset Lewy bodies absent Good response to Levodopa Slow progression	Various point mutations Exonic rearrangements	≥30	Kitada et al. 1998 Lucking et al. 2000 Brüggemann, and Klein, 2013
***PINK1* (PARK6)**	Early-onset Lewy bodies absent Good response to Levodopa Slow progression	Various point mutations Exonic rearrangements	≤50	Valente et al 2004 Cazeneuve et al. 2009 Li et al. 2005
***SYNJ1* (PARK20)**	Early-onset Cognitive decline Fast progression	p.R258Q p.A320V p.A551V p.S1422R	≤50	Krebs CE et al. 2013 (Quadri et al., 2013) (Drouet and Lesage, 2014) Chen et al. 2015

Several other genes associated with PD, have been identified (**Appendix I, Table 1**). However, most of these genes have only been found in a single family and to date there is no reported experimental studies on the impact of protein dysfunction leading to PD pathology. Further investigation is needed to elucidate their role in PD pathogenicity or their contribution to disease risk (Golbe et al., 2001).

1.3. Whole Exome Sequencing Technology and its Contribution to Parkinson's Disease Genetic Research

1.3.1. Next Generation Sequencing Technology and Whole Exome Sequencing

The modern history of DNA sequencing technology began in 1977, when two methods namely chemical cleavage DNA sequencing and chain-terminating DNA sequencing were introduced (Maxam and Gilbert, 1977; Sanger et al., 1977). These first-generation sequencing technologies relied on the separation of DNA fragments of various sizes on polyacrylamide gels to determine the sequence of nucleic acids. Over time the Gilbert and Maxam method was replaced by Sanger sequencing because it was less complex and involved the use of less hazardous chemicals. The Sanger sequencing method has remained the "gold standard" of DNA sequencing for decades and was the core technology of the Human Genome Project. This project aimed to determine all three billion base pairs making up the human genome (International Human Genome Sequencing Consortium, 2001). In 2001, after 15 years of effort the first human genome was sequenced by the Human Genome Project at an estimated cost of $2,7 billion (Lander et al., 2001; Venter et al., 2001).

Over the years, DNA sequencing has advanced with the development of parallel sequencing high-throughput technology known as next generation sequencing (NGS). It is now a large-scale industry with specialised infrastructure of robotics, instruments, bioinformatics and computer databases. These technologies are different to Sanger sequencing where one run only produces a single forward or reverse sequencing read, whereas NGS allows millions of sequencing reactions to be carried out in parallel (Baudhuin et al., 2015). NGS presently enables researchers to sequence the entire human genome in under two weeks for approximately US $ 1506,00 (**Figure 1.4**) and the entire human exome (protein coding regions of DNA) of an individual for US $555,00 in under 48 hours. This represents remarkable progress in reducing the cost and time of DNA sequencing by several orders of magnitude relative to first-generation sequencing technologies (Schwarze et al., 2018).

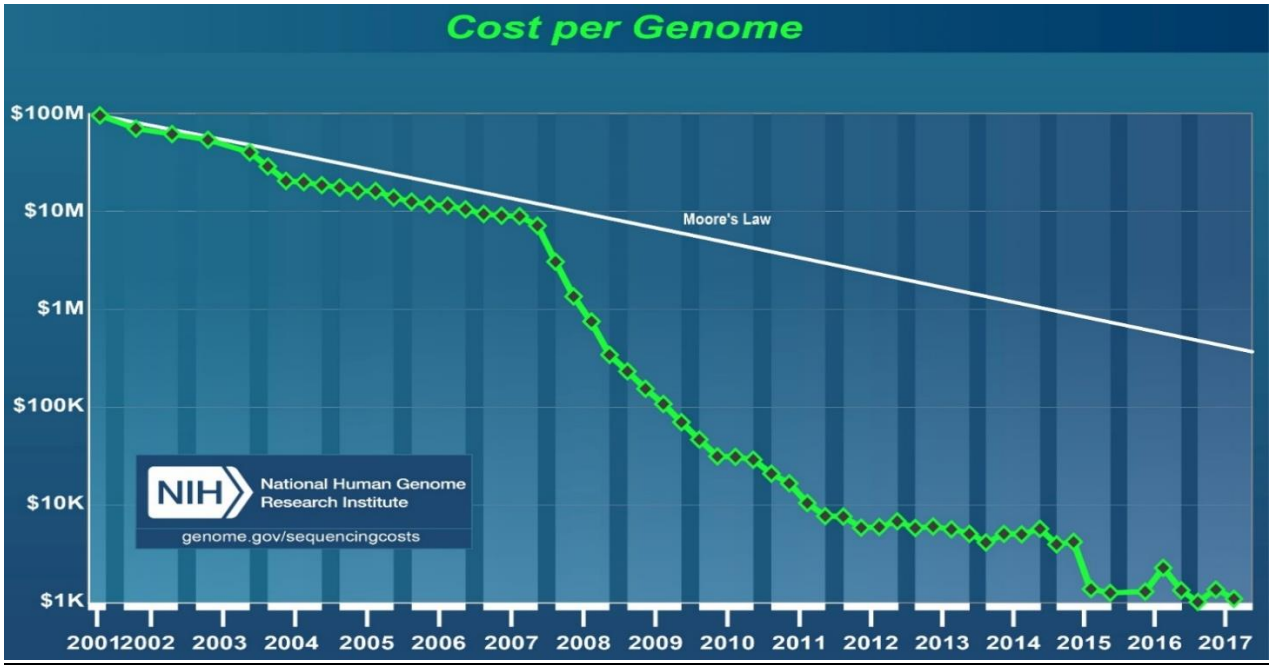


Figure 1.4 Cost of sequencing a human genome over a 17-year period. Since 2001 the National Human Genome Research Institute (NHGRI) has tracked the costs associated with DNA sequencing of human-sized genomes. On the y-axis is the cost of sequencing in US Dollars and on the x-axis is time in years. The sudden and profound drop in cost observed between 2007 and 2008 represents the time when most sequencing centres transitioned from first-generation sequencing to NGS technologies (Used with permission from the NHGRI https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/).

The NGS workflow is typically made up of five main steps (**Figure 1.5**), first the sample collection phase where DNA isolation and purification from the samples of interest is performed. The DNA (for humans) may originate from a variety of sources and ideally between 10 and 20 µg of DNA is needed to produce a library (Mardis, 2008; Metzker, 2010). The second phase is library construction, where the DNA template is converted into a library of sequencing reaction templates. Each sequencing platform uses a slightly modified library preparation process, but in essence the DNA is randomly sheared into specific sizes suitable for sequencing by sonification, nebulisation or enzyme digestion (Mardis, 2008).

The resulting DNA fragments are then ligated to platform specific adaptor oligos at both ends. If the sequencing is targeted at a specific region of the genome like the exome or at a certain group of genes, then enrichment of the target fragments for that region is done. Enrichment can be done using variable capture strategies such as hybrid capture, microdroplet PCR, or array capture techniques (Schuster, 2008; Mardis, 2008).

The third phase of NGS is the library immobilization phase. This involves the DNA templates being separated into single strands and then attached by the already bound

oligo onto beads (Emulsion PCR) or a solid surface (Bridge PCR) before extension and amplification (Schuster, 2008; Mardis, 2008). This immobilization of the template strands allows for the millions of reactions to be carried out in parallel on each template.

The fourth step of the NGS workflow is sequencing, this step greatly varies according to the platforms sequencing chemistry. The most commonly used sequencing chemistries are cyclic reversible termination, real-time sequencing, pyrosequencing, semiconductor sequencing and sequencing by ligation. This step results in the sequence of each template strand being read out after the incorporation of each base through fluorescence or chemiluminescence (Metzker, 2010).

The final step of the NGS work flow is the data analysis, where bioinformatics programs convert the generated raw output to interpretable information. The detected read data is aligned to a reference or assembled *de novo*, then primary filtering removes duplicate reads and those that have low quality scores. Lastly single nucleotide polymorphisms (SNPs) and inversion-deletions can be identified before downstream interpretation of their impact (Mardis, 2008).

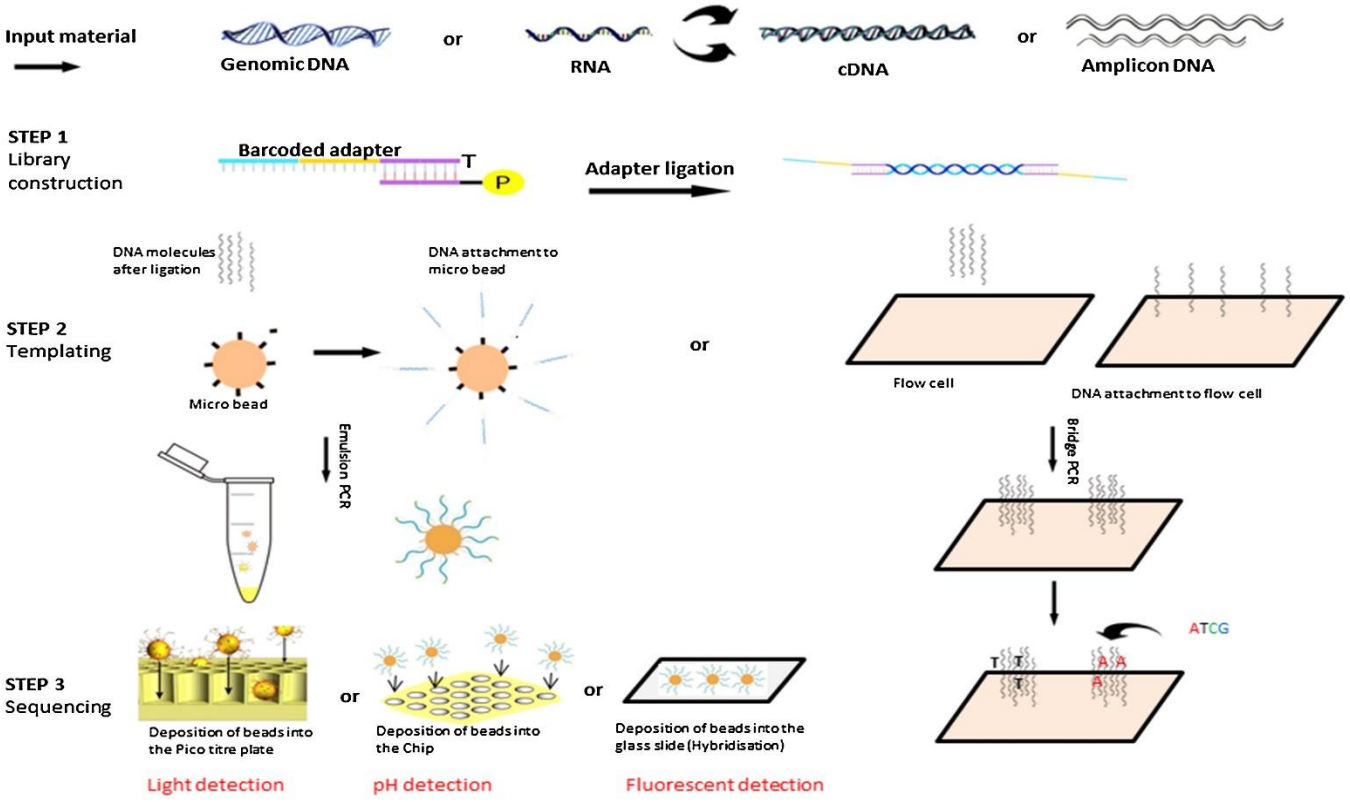


Figure 1.5 Schematic representation of the NGS workflow. Sample collection, library construction, template immobilization, sequencing and detection. After the signal detection step, data interpretation begins with alignment, mapping, filtering and SNP calling using bioinformatic computational tools (Pillai et al., 2017) . Used with permission from Elsevier.

Whole genome sequencing (WGS), the most comprehensive NGS method, sequences the non-coding regions of the genome in addition to the exons. It has already proven its usefulness however, the large output of data obtained and the complexity in interpreting it remains a computationally intensive challenge (Meynert et al., 2014). WES another NGS approach, extracts only the protein coding fragments of the genome. The commercially available WES kits target all of the genes documented in Refseq (Pruitt et al., 2014). This typically will yield between 20,000 to 25,000 variants per individual (Bamshad et al., 2011).This constitutes approximately ~1,2% of the genome which provides the advantage of producing more manageable sequencing data output than WGS (Bamshad et al., 2011).

1.3.2. Advantages of Whole Exome Sequencing

The greatest benefit of WES has been the exponential increase in the rate of discovery of Mendelian disease genes (Bamshad et al., 2011). Previously, screening for mutations was conducted in known regions of interest. Currently, exomes which are predicted to harbour ~85% of the mutations, can all be simultaneously screened (Botstein and Risch, 2003). This has been of great benefit in genetically heterogeneous conditions (single disease being caused by any one of a multiple number of alleles) like PD. It is cost effective and quicker than both Sanger sequencing and targeted gene panels, which lag behind in integrating newly associated disease genes (Ng et al., 2010). Apart from gene discovery, WES has also been useful in genetic diagnostic settings. Using WES Yang et al., (2013) correctly diagnosed 62 out of 250 children with neurological phenotypes, which is a higher diagnostic yield than the other genetic tests (karyotype, chromosomal microarray and Sanger sequencing) in a fraction of the time (Yang et al., 2013).

Compared to WGS, WES yields greater sequencing accuracy in a fraction of the time. The amount of data generated from WES at 100X depth is ±5 gigabyte which is significantly less than WGS at the same depth yielding ±90 gigabyte (Cherukuri et al., 2015). Therefore, the data storage costs of WES are lower, data analysis is rapid and more manageable. For example, it is currently possible to sequence the exomes of three different individuals at a target coverage of 40X for the same cost as a single human genome at 30X coverage (Robert et al., 2014).

1.3.3. Disadvantages of Whole Exome Sequencing

In spite of the many successes, there are technological limitations to WES. One disadvantage is its inability to accurately detect structural variants such as inversions and translocations, as well as other types of variants such as non-coding variants, indels,

copy number variations and repeat expansions (Botstein and Risch, 2003; Krumm et al., 2012). As the field is advancing this can be to some degree overcome by a number of tools that have been proposed in the literature, that detect structural variant signatures which are diagnostic of different rearrangements with high sensitivity and specificity when used in combination (Tattini et al., 2015). Some non-coding variants may be overlooked despite their importance for controlling gene transcriptional regulation or splicing (Fromer et al., 2012). Although WES yields on average a higher coverage than WGS, the target coverage along the length of exons in WES is usually uneven. This causes errors in downstream analysis such as false-positive calls in variant calling analysis. Factors that affect coverage in WES include the size of the exon, the presence or absence of repeat elements, GC content and segmental duplications (Kozarewa et al., 2009; Belkadi et al., 2015). Other errors can occur during WES due to hybridisation not being stringent enough to differentiate between exons belonging either to genes with very similar sequences such as paralogues, pseudogenes, and gene family members (Gnirke et al., 2009).

1.3.4. Computational Tools and Variant Filtration Approaches

The approach of filtering through NGS data typically consists of the following steps: raw data quality assessment, alignment, variant calling, annotation, and then prioritisation (Bao et al., 2014). The quality of the raw data is first assessed using several tools such as FastQC (Andrews, 2016) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), FastQ (http://homer.ucsd.edu/homer/basicTutorial/fastqFiles.html) and FASTX-Toolkit (Gordon and Hannon, 2010) (http://hannonlab.cshl.edu/fastx_toolkit/) to determine whether pre-processing steps such as base trimming, read filtering, or adaptor clipping are necessary. The sequences are then mapped/aligned, and this entails the cross-referencing of the 20,000–25,000 SNPs to the reference genome using tools such as Sequence Alignment/Map (SAM) (Li et al., 2009) (http://software.broadinstitute.org/software/igv/sam), Binary Alignment/Map (BAM) (Li et al., 2009) (https://software.broadinstitute.org/software/igv/BAM), Bowtie2 (Langmead and Salzberg, 2012) (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) and Novoalign (http://www.novocraft.com/products/novoalign/).

Variant calling involves producing meta-information for each variant relative to the reference genome sequence, as well as quality metrics, like the overall number of reads at each position and the number and depth of alleles detected (Liu et al., 2013). Popular variant calling tools include SAMtools (Li et al., 2009) (https://samtools.github.io), GATK (McKenna et al., 2010) (https://software.broadinstitute.org/gatk/) and Atlas2 (Challis et

al., 2012) (https://www.hgsc.bcm.edu/software/atlas-2). After variants are called, annotation adds attributes such as genomic feature, gene symbol, exonic number, amino acid change and the effect of the variant on the gene product (non-sense, missense, splice-site SNPs, insertion, deletion or inversion, etc.) (Rabbani et al., 2014).

Prioritisation is the part of filtration that reduces the total number variants to a workable list of candidates. This process is versatile and varies across studies according to the methods and tools that are most informative and relevant. The ultimate purpose of prioritising, regardless of the combination of tools, is to remove any unreliable variant calls, to determine the frequency of variants in various populations and to assign priority to variants most relevant to that specific disease (Bao et al., 2014).

A common step in filtration is to determine the population frequency of the variants. It is expected that Mendelian disease-causing alleles are extremely rare, if not absent, in the general population due to negative selection (Wong et al., 2003). The minor allele frequency (MAF) is the frequency of the less frequent allele at a given locus and in a given population. In AD alleles, the MAF would be expected to be low as the variants would not be commonly found in the general population, variants would be either novel or extremely rare. While the alleles responsible for causing AR diseases can be found in the general population (Wong et al., 2003). The MAF threshold of ≤1% is typically used to filter out polymorphisms.

Publicly available online databases can be used to determine the frequency of the variant in different populations. The commonly used databases include the Exome Aggregation Consortium (ExAC) Database (Karczewski et al., 2017), the 1000 Genomes Project (Siva, 2008, The 1000 Genomes Project Consortium, 2015), dbSNP (Sherry et al., 2001) and NHLBI Exome Sequencing Project (https://esp.gs.washington.edu/drupal/). Populations represented in these databases include, but are not limited to, the European (Non-Finnish), African, East Asian, Latino and South Asian populations. The MAF threshold of ≤1% though useful, should be varied depending on the frequency of the disease in the population of interest and according to the expected penetrance of the candidate variant. Since the frequency of benign and pathogenic variants can differ between ethnicities, the variant might be absent in one population but commonly found in another (Wong et al., 2003). Therefore, when determining the population frequency of variants, it is essential to also screen ethnically-matched populations.

The degree of amino acid conservation can also be used to prioritise variants. Conservation prediction programs like GERP (Davydov et al., 2010), PhastCons (Siepel, 2005) and

PhyloP (Pollard et al., 2010) determine conservation by aligning human protein sequences to homologous protein sequences from other organisms. The rationale behind this is that the more conserved a region is within the multiple alignment, the more damaging an amino acid-changing variant at that position will be. Alternatively the less conserved the region is predicted to be, the more likely change will be tolerated (Davydov et al., 2010).

The final step of variant prioritisation is usually the predicting of effects of protein coding sequence changes. To do this bioinformatic *in-silico* approaches have been developed to predict the functional impact of a SNP on the protein's structure and function (Patel et al., 2014). Each tool is unique in incorporating different information to make its prediction this includes evolutionary conservation, biochemical properties, statistical significance and what is commonly known from literature.

1.3.5. Bioinformatic *In-silico* Tools to Assess Pathogenicity

Numerous tools, capable of classifying both non-synonymous and structural variants are publicly available such as Provean (Protein Variation Effect Analyser) (Choi and Chan, 2015), CADD (Kircher et al., 2014), FATHMM-MLK (Functional Analysis through Hidden Markov Models) (Shihab et al., 2015), GWAVA (Genome Wide Annotation of VAriants) (Ritchie et al., 2014) and MutationTaster (Schwarz et al., 2014).

The most popularly used tools for pathogenicity prediction since their advent have been SIFT (Sorts Intolerant From Tolerant) (Ng and Henikoff, 2003) and PolyPhen-2 (polymorphism phenotyping version 2) (Adzhubei et al., 2013). Both tools give probabilistic estimates based on conservation of protein sequence between species. Polyphen-2 additionally incorporates the physicochemical features of amino acids, and structural features of protein to generate the predictions in a Bayesian fashion. These tools differ not just in their inference algorithms but also in composition of predictive features which explains their occasional contrasting predictions (Flanagan et al., 2010). Comparable performance tests can be found in their original reports (Ng and Henikoff, 2003;Kumar et al., 2009; Adzhubei et al., 2013), as well as in recent independent assessment studies (Li et al., 2013; Ghosh et al., 2017). Nonetheless, in terms of overall performance across the independent assessments conducted using known mutations, SIFT and PolyPhen-2 fare similarly on prediction accuracy.

New meta-tools have been proposed that appear to achieve better performance, by combining prediction scores from multiple tools. CADD uses a unified score for the potential deleteriousness of all 8,6 billion possible human mutations by comparing variants that survived natural selection, with simulated mutations. The tool integrates

63 predictive features in total, including SIFT, GERP++, PolyPhen-2 scores, as well as CPG distance and GC content. It measures deleteriousness, a property said to encompass both molecular functionality and pathogenicity, compensating for the incompleteness and bias of the existing predictive tools (Kircher et al., 2014).

Multiple studies have assessed the utility of CADD to prioritise functional and disease relevant mutations. One such study was conducted by Bandaru et al., (2015) on previously been annotated and functionally validated WES data from a male of Chinese descent. As the variants impact outcomes were known the CADD C-scores predictions could be compared to the existing prediction tools for accuracy. The CADD C-score and predictions outperformed both the missense-only predictive tools and the conservation tools in predicting variant impact. Another study by Mather et al., (2016) had conducted germline cancer susceptibility testing on 624 patient samples that had generated 12,391 novel SNPs. Although accurate in the prediction of the missense variants, they concluded that the algorithm calculating the CADD score was limited in magnitude to effectively classify the significance of non-exonic variants. This sentiment had been previously expressed for the existing predictive tools and therefore none of them should be independently used to prioritise candidates (exonic and intronic) (Mather et al., 2016).

Although they are greatly useful, functional prediction tools should not be the sole method of prioritisation, as each tool is based on certain assumptions and therefore have their own limitations. The groups that have produced these tools have emphasised that the terms damaging or deleterious should never be logically equated with causal for a disease phenotype. As a variant that damages a gene is not necessarily damaging to the overall health of an individual. These tools merely seek to improve filtration accuracy and should be used in combination where possible and in conjunction with additional sources of information such as knowledge about the molecular and cellular mechanisms involved in a disease or phenotype.

1.3.6. Whole Exome Sequencing Approaches to Identify Parkinson's Disease-causing Genes

Before 2011, many of the earlier genetic studies on PD involved genetic linkage analysis which brought about the discovery of *SNCA, LRRK2, PARK2, PINK1* and *DJ-1* (Polymeropoulos et al., 1997; Kitada et al., 1998; Bonifati et al., 2003; Valente et al., 2004; Paisán-Ruíz et al., 2004). However, this approach was labour intensive and expensive due to the large number of genes to be sequenced by Sanger sequencing. It also requires large pedigrees with both affected and unaffected individuals. WES presented the opportunity

to investigate all the known genes simultaneously. This has contributed to the discovery of new PD associated genes *CHCHD2* (Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 2), *VPS35* (Vascular Protein Sorting 35) and *LRP10* (Low-density lipoprotein receptor-related protein 10) all found in familial PD (Zimprich et al., 2011; Vilariño-Güell et al., 2011; Funayama et al., 2015; Quadri et al., 2018).

The first studies that used WES to discover the PD associated gene *VPS35* were conducted by Vilariño-Güell et al., (2011) and Zimprich et al., (2011). The study by Vilariño-Güell et al., (2011) was conducted on a multi-incident Swiss family with AD, late-onset PD. The affected family members had a mean age at onset of 50,6 ± 7,3 years and were tremor-predominant with a good response to LD. They had been previously screened for the known PD mutations which were absent, and WES was applied to interrogate co-segregating variants amongst the affected family members (Vilariño-Güell et al., 2011), identified 4,256 novel variants.

Authors assumed AD inheritance of disease, and therefore excluded variants on the X-chromosome and homozygous changes. Removal of the non-coding variants, synonymous changes, and variants present in dbSNP identified 69 variants. Sanger sequencing of the 69 variants in one previously whole exome sequenced affected member, two additional affected family members and one unaffected family member, validated 33 variants, six of which co-segregated exclusively in family members with PD. Screening 184 randomly US control subjects of European descent for the six variants revealed that only two were not in the control population, namely one within the *VPS35* gene and one in Integrin Subunit Alpha X (*ITGAX*).

Both variants were subsequently genotyped in a multi-ethnic case-control series consisting of 4,326 patients and 3,309 controls. Nine affected individuals and none of the controls had the *VPS35* Asp620Asn (c.1858G>A) variant. The *VPS35* gene was found to be highly conserved and the p.Asp620Asn substitution was predicted to be a deleterious substitution (SIFT score = 0). Together, this information implicated the *VPS35* Asp620Asn (c.1858G>A) variant in being a strong candidate variant for PD. The second variant *ITGAX* p.Ala1012Val mutation was found in one control (61 years) and one affected individual (42 years at onset). The *ITGAX* gene was predicted to not be evolutionarily conserved, and p.Ala1012Val was not predicted to be deleterious (SIFT score = 0,16), and was authors concluded that *ITGAX* p.Ala1012Val was a rare variant (Vilariño-Güell et al., 2011).

Another study by Zimprich et al., (2011) performed WES on an Austrian family with 16 affected individuals with late-onset PD. They discovered the p. Asp620Asn (c.1858G>A) mutation in *VPS35* which co-segregated in seven out of the 16 affected family members. This variant was not found in the 860 sporadic PD cases and in the 1014 controls. Two additional families, with five and ten affected members with late-onset PD, were further screened and the p. Asp620Asn mutation in *VPS35* co-segregated in both families with incomplete penetrance in an AD fashion. Additionally, sequence-based and molecular-dynamics based predictive tools PolyPhen2, SNAP and SIFT were utilised and all the tools predicted the *VPS35* p. Asp620Asn to be deleterious. A second variant in *VPS35* was detected in a single sporadic PD case p.Arg524Trp (c.1570C>T) and was also predicted to be damaging by sequence-based and molecular-dynamics analyses (Zimprich et al., 2011). *VPS35* is a protein critical for endosome-trans-Golgi trafficking and membrane-protein recycling (Vilariño-Güell et al., 2011).

A study by Funayama et al., (2015) performed WES on a Japanese family with eight affected members who had an average age of onset of 55,5 ± 4,8 onset with AD PD. Four of the eight affected individuals were WES and a heterozygous variant p.T61I (c.C182T) in *CHCHD2* was identified. The variant co-segregated variant in all eight affected individuals and was not present in five unaffected family members. The mutation was further screened and found in two out of 340 familial PD patients but not found in 517 sporadic PD patients or 559 ethnically-matched controls. The variant was not found in online population databases 1000 genome and Human Genetic Variation Database (Funayama et al., 2015). *CHCHD2* is known to regulate mitochondrial crista structure and maintaining the integrity of the mitochondrial respiratory complexes. The dysfunction of the gene has been shown to impair mitochondrial function significantly, this could be the mechanism by which it contributes to the PD pathology (Meng et al., 2017) .

A recent study used genome wide linkage analysis in an Italian family with ten affected individuals with an average age at onset of 59,8 years (± 8,7) to identify a novel locus on chromosome 14 (Quadri et al., 2018). WES was then performed on the index patient to identify candidate variants within that specific locus. The resulting variants were then filtered based on being non-synonymous variants in a heterozygous state, having a MAF of less than ·1% in four online population frequency databases and lastly being predicted as pathogenic by at least five of 11 in-silico tools. The prioritised candidate *LRP10* p.G603R (c.G1807A) was present in all the affected individuals, absent the unaffected

family members and eight of the 11 functional prediction tools classified it as deleterious. All the exons of *LRP10* were then screened for variants in 938 unrelated PD patients from seven different countries. This yielded an additional seven rare and potentially pathogenic *LRP10* variants (p.G603R, p.P699S, p.Y307N, p.R533L, p.R235C, p.1424+5delG, p.N517del and n.A212Sfs17), present in multiple affected PD patients. Control screening was performed on 553 unrelated controls and 92 relatives of PD patients, using WES. Only one individual out of the 645 screened controls was found to have a variant (p.R151C) in *LRP10* (Quadri et al., 2018). LRP10 has been shown to function as a shuttle between the Golgi network, endosomes, and plasma membrane similar and in pathways related to PD gene *VPS35* (Steinberg et al., 2013).

1.4. History of Molecular Research of Parkinson's Disease in South Africa

In comparison to the global advances in understanding the genetics of PD, studies on SSA populations have been scarce. The majority of genetic PD research in this region was conducted in SA and has focussed on candidate gene studies (all studies found summarised in **Appendix I, Table 2)**. Autosomal dominant PD genes *EIF4G1, GBA, GCH1, LRRK2, SNCA, VPS35* and autosomal recessive PD genes *ATP13A2, DJ-1, Parkin* and *PINK1* have been assessed in SA PD patients using various assays, including KASP genotyping assay, dosage analysis MLPA (Multiplex Ligation-dependent Probe Amplification), HRM and Sanger sequencing.

Of the five studies that screened for *LRRK2* mutations in SA PD patients, only two studies identified SNPs (Bardien et al., 2010), (Keyser et al., 2010), (Blanckenberg et al., 2014), (Mahne et al., 2016), (Barkhuizen et al., 2017). The first study, involving 205 PD patients, found that four patients harbour the G2019S mutation in a heterozygous state (Bardien et al., 2010). Another study conducted by Trinh et al., (2015) screened all 51 exons of *LRRK2* in four SA PD patients, and identified four rare non-synonymous variants that were unique to this population. Keyser et al. (2009), investigated *SNCA* in 88 patients, and only one affected individual was found to have three copies of the gene (Keyser et al., 2009). None of the other AD mutations in known PD genes were found by various studies that were conducted on the SA population. (Keyser et al., 2010, Blanckenberg et al., 2014, Barkhuizen et al., 2017),.

A study screening exon two and exon nine of *ATP13A2* in 88 SA PD patients yielded none of the known mutations (Keyser et al., 2009). In *DJ-1,* a novel 16 bp deletion variant (g.-6_+10del) was found in the 5' untranslated region of the gene by two separate studies. The first study identified the g.-6_+10del variant in two out of 150 SA PD patients, however

screening of an additional 30 PD patients did not identify other individuals with this variant (Keyser et al., 2009). Subsequent screening of a larger cohort (402 SA individuals) for the g.-6_+10del variant identified in the same individual as the Keyser et al. (2009) study, representing an allele frequency of 0,1 % (1/804). This study also found another deletion g.168_185del present with an allele frequency of 0,6 % (5/804) using the KASP genotyping assay (Glanzmann et al., 2013). Variant g.168_185del was proposed to interfere in the role of DJ-1 in alleviating oxidative stress, however because it was present in both cases and controls, it was concluded that it was unlikely that it had a significant role in the development of PD in this population (Glanzmann et al., 2013). A luciferase assay was used to show that the g.-6_+10del variant reduces DJ-1 expression in the response to oxidative insults (Keyser et al., 2009).

Mutations in the coding regions of the Parkin gene are the most common cause of early-onset PD cases. These include heterozygous and homozygous point mutations, non-sense mutations and exonic deletions (Myhre et al., 2008). In SA PD patients, known mutations in *PARK2* were found in four separate studies, in both heterozygous and homozygous forms (Keyser et al., 2009, Haylett et al., 2012, Mahne et al., 2016 ,Carr et al., 2016). A WES approach in an affected family identified two mutations, namely p.R275W and p.M432V, in homozygous forms in three PD patients, and in heterozygous form in three unaffected family members (Carr et al., 2016). Population screening on the ExAC database showed that the R275W variant had an allele frequency of 0,2%, whereas the M432V variant's frequency is 0,00096%. Bioinformatic tools PolyPhen2, MutationTaster and FATHMM predicted M432V as pathogenic with a CADD score of 20,6 (Carr et al., 2016). A separate study using MLPA and real-time PCR techniques identified the polymorphism p.M192L in 13 out of the 210 affected individuals (van der Merwe et al., 2016). The study also found that one patient out of the 210 had a heterozygous *PARK2* exon 4 deletion. They concluded that the *PARK2* mutations previously discovered in other populations were not a major contributor to PD pathology in the SA population (van der Merwe et al., 2016).

The screening for *PINK1* variants in 154 SA PD patients using the HRM curve analysis and direct sequencing yielded a total of 16 sequence variants (Keyser et al., 2010). Thirteen of these were previously known polymorphisms, while five were novel. Two of the three remaining variants (p.Y258X and p.E476K) had been previously reported (Valente et al., 2004; Rogaeva et al., 2004; Bonifati et al., 2005; Abou-Sleiman et al., 2006; Keyser et al., 2009; Tan et al., 2006). The p.Y258X mutation was found in homozygous

form, and p.P305A and p.E476K were in heterozygous form (Keyser et al., 2010). The p.P305A variant, which was novel, was found in a black female patient who had a positive family history for PD with an age at onset of 30 years. The variant was predicted to be probably damaging by PolyPhen and was found to have an allele frequency of 1,9% (2/108) amongst the controls. Structural protein analysis showed the potential to modulate enzymatic activity due to its location in the kinase domain. (Keyser et al., 2010). Two other CNV studies that also screened the eight exons of *PINK1* using MLPA and CNV but did not find any pathogenic mutations (Keyser et al., 2009, van der Merwe et al., 2016).

A study conducted by van der Merwe et al., (2016) aimed to determine the CNV of multiple PD genes on 210 SA PD patients using MLPA (van der Merwe et al., 2016). The genes screened were *SNCA*, *PINK1*, *LRRK2* (exons 1, 2, 8, 10, 15, 27, 49) and DJ-1 (exons 1, 3,5, 7). The study calculated the combined frequency of exonic rearrangements in the known PD genes by including results from a previous study on 229 SA PD patients. The frequency of exonic rearrangements was reported to be 1,8% (8/439 patients) in SA PD patients (Keyser et al., 2009; van der Merwe et al., 2016; Haylett et al., 2012). This means that although CNV has been implicated in PD pathogenesis in other populations, in SA patients it is very rare and not a major contributing factor to PD pathogenesis.

Lastly a large collaborative study performed by the Consortium of Genetic Epidemiology of Parkinson's Disease (http://geopd.can.ubc.ca) combined PD patient populations from multiple countries, including 398 unrelated SA PD patients. These individuals were screened for polyglutamine repeat expansions in the Spinocerebellar Ataxia genes, but the study did not find any pathogenic repeat expansions in *CACNA1A, TBP, ATXN2* or *ATXN3* in the SA PD patients (Wang et al., 2015).

1.5. The Present Study

There is a clear disparity in the number of PD genetic studies conducted on SSA populations, in comparison to the studies conducted in Caucasian and Asian populations. This is thought to be due to extrinsic factors such as the lack of resources and skilled personnel, and the shorter life expectancy of SSA populations (Okubadejo, 2006, Blanckenberg et al., 2013). However, as the number of older people in SSA is growing faster than the rest of the world, there is an increasing need to study age-related diseases like PD in the SSA population (Velkoff et al., 2007).

It is well established that the known PD mutations are minor contributors to disease aetiology in SSA populations (Bardien et al., 2009; Blanckenberg et al., 2013; Cilia et al.,

2011; Glanzmann et al., 2013; Haylett et al., 2012; Keyser et al., 2009a, 2009b, 2010b; Okubadejo et al., 2008; Okubadejo, 2008; van der Merwe et al., 2016; Wang et al., 2015; Yonova-Doing et al., 2012). This further highlights the need to bridge the knowledge gap in the genetics of PD in SSA, by using new technologies to discover novel PD-causing genes or novel mutations in the known genes. Ultimately, this knowledge could make an important contribution to the current understanding of PD genetic aetiology both locally and globally.

The present study investigated the potential genetic cause of PD in a SA Afrikaner family with multiple affected individuals. Given the uniqueness of this ethnic group, we undertook WES as an approach to identify a possible novel PD-causing gene in the family.

The Afrikaner Caucasian population is unique to SA (Greeff, 2007) and is derived from settlers that arrived in SA thought to have mainly originated from the Netherlands, Germany and France (Heese, 1971). Initially, the settlers lived in isolated communities in relatively small communities, with very little admixture with the surrounding populations. This is thought to have contributed to the relatively high frequency of founder effects in this ethnic group. Some of the common Mendelian disorders in this population can even be traced back to specific founder couples. This includes progressive familial heart block (Brink et al., 1977), Huntington's disease (Hayden et al., 1980) osteogenesis imperfecta (Knoll et al., 1988), pseudoxanthoma elasticum (Saux et al., 2002), schizophrenia (Karayiorgou et al., 2004), long QT syndrome (Brink et al., 2005) and Fanconi's anaemia (Tipping et al., 2001).

Notably, in a previous study conducted by our research group genealogical analysis revealed that six unrelated Afrikaner PD patients were related to a single founder couple that immigrated to SA in the 1600s, suggesting a possible founder effect for the disease (Glanzmann, 2013; Geldenhuys et al., 2014). The same study also aimed to identify novel PD-causing gene using WES, 21 variants were found to be shared by three unrelated Afrikaner PD patients (Glanzmann, 2013). This present study is different in that we will investigate the cause of PD in a single Afrikaner family. It is anticipated that studying a familial form of PD will increase the likelihood of discovering a novel PD-causing gene.

1.6. Aims and Objectives

The aim of this study was to investigate the genetic aetiology underlying PD in a multiplex SA Afrikaner family. The aim was achieved through the following objectives:

i. To perform WES on three PD affected family members.

ii. To apply a comprehensive filtration strategy to prioritise only co-segregating variants (using bioinformatic tools) found in the three PD patients.

iii. To validate the presence of variants discovered with WES and screen the unaffected family members for the variants by Sanger sequencing.

iv. To perform high resolution melt (HRM) analysis to determine the frequency of the variants in ethnically-matched controls.

v. To perform functional studies on a single prioritised candidate variant, to assess its potential functional effect.

Chapter Two
2. Materials and Methods
A. Whole Exome Sequencing Analysis and Prioritisation

In this section we use WES to screen a multiplex family with PD, then validate the list of variants using Sanger sequencing. Computational analysis is then performed on the validated variants to predict the pathogenicity before prioritising the list of candidates to a single gene.

2.1. Participant Recruitment

2.1.1. Ethical Considerations

Ethical approval was obtained from the Health Research Ethics Committee of Stellenbosch University, Cape Town, SA (Protocol number 2002/C059) (see **Appendix II, Figure 2**). Written informed consent was obtained from all study participants.

2.1.2. Study Participants and Selection Criteria

In the context of this study, the term 'Afrikaner' refers to the Afrikaner Caucasian population that is unique to SA. The selected family (labelled as ZA 253) was of Afrikaner ethnicity and consisted of four affected PD patients and five unaffected individuals .The five affected patients were examined by a specialist and diagnosed according to the UK Parkinson's Disease Society Brain Bank Research criteria (Gibb and Lees, 1988). The age at onset (AAO), lifestyle information and family history were obtained from the patients by a trained research nurse using a questionnaire. The average AAO of the affected individuals was 50,8 years (ranging from 37 to 77 years).

A total of 192 healthy, unrelated individuals were recruited as controls though these individuals were not clinically assessed for PD. Of the 192 controls, 96 were self-identified as SA Afrikaner Caucasians and 96 were self-identified as SA 'non-Afrikaner' Caucasians. Males made up 51,6% of the controls, and the average age was 64,5 (±19,9) years.

2.1.3. Extraction of gDNA from whole blood

Peripheral blood samples were collected from the study participants and genomic DNA (gDNA) was extracted from whole blood using the NucleoSpin® Blood XL kit (Macherey-Nagel™ GmbH and Co., Germany), according to manufacturer instructions.

In brief, 10 ml of whole blood is mixed with 500µl of Proteinase K and 10 mL Buffer BQ1, which lyses the cells. The sample enters a mixing and incubation cycle where it is vortexed for 10 seconds and then it is incubated at 56 ˚C for 10 minutes, this occurs three times

before cooling the lysate to room temperature. Once cooled, 10 ml of Ethanol (96–100 %) is added before mixing by inversion allowing the DNA to precipitate. The 15 ml of lysate is then transferred to a collection tube with a NucleoSpin® Blood XL Column. The sealed collection tube is then centrifuged 3000 rpm for 3 minutes and the flow through is discarded. This is repeated with the remaining 15 ml lysate in the same column. Once the second flow through has been discarded the column still in the collection tube is washed with 7.5 ml of Buffer BQ2 and centrifuged for two minutes at 3000 rpm. The wash step is repeated however it is now centrifuged for 10 minutes at 3000 rpm. After placing the washed column in a new collection tube, the DNA is then dissolved by pouring 1000 µl of preheated Buffer BE (70 °C) directly onto the membrane and centrifuging at 3000 rpm for two min. The elution collected contains an ultra-pure, high-yield DNA which will be stored at 4°C, ready to use for downstream analysis. DNA concentration was estimated by measuring the absorbance at 260nm on the NanoDrop spectrophotometer.

2.2. Whole Exome Sequencing

At the time when the sequencing was performed only three of the five affected individuals were analysed (12.726, 11.844 and 92.32) because individuals 92.33 and 11.902 had not yet been diagnosed with PD. WES was performed using an Illumina HiSeq 2000 platform by a collaborator, Prof. Matt Farrer (University of British Columbia, Canada).

Genomic DNA samples were constructed into Illumina PairEnd pre-capture libraries according to the Illumina Multiplexing Sample Prep Guide 1005361 protocol (https://support.illumina.com/downloads/multiplexing_sample_prep_guide_1005361.html). Library templates were prepared for sequencing using Illumina's cBot cluster generation system with the corresponding TruSeq PE Cluster Kits (Illumina© Inc, United States) for the HiSeq. Briefly, the library was denatured with sodium hydroxide and diluted to 3 to 6 pM in hybridisation buffer in order to achieve a load density of 700 to 900 k clusters/mm2 on the HiSeq 2000. Each sample was pooled (in a set of three) for loading onto a single lane of a HiSeq flowcell and then deconvoluted for analysis based on barcode sequence. All lanes were spiked with 2% phage phiX DNA control library for run quality control. After loading onto the flow cell, sample libraries underwent bridge amplification to form clonal clusters, and the sequencing primer was hybridised.

Sequencing runs were performed in paired-end mode using the HiSeq 2000 platform. Using the TruSeq SBS Kit, sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional seven cycles for the index read for runs that included pooled samples. Real-time analysis software was used to process the image

analysis and base calling. Sequencing runs generated approximately 100 to 140 million successful reads per lane. With these run yields, each sample library achieved 10 to 12 Gb of raw DNA sequence data, which enabled a minimum of 20x coverage for 85% to 90% of the bases targeted in the exome.

Burrows-Wheeler Aligner (BWA-MEM) (Li, 2013) was used to align sequence reads or assembly contigs against the Genome Reference Consortium Human NCBI build 37 (GRCh37/hg19). BWA-MEM was also used for the removal of read duplicates, SNP calling and indel detection. ANNOVAR Software (Wang et al., 2010) was used to annotate the variants using GRCh37/hg19 as the reference genome.

The Genome Analysis Toolkit (GATK) was used for variant calling, and to describe variants as either exonic, intronic, in the UTR or splice-site region by comparison with a reference sequence (Wang, Li, and Hakonarson 2010). To provide a summary of the coverage of mapped reads on a reference sequence at a single base pair resolution SAMtools mpileup was used with a Minimum variant QUAL score of 30. The co-segregating variants of the three PD patients that met the filtration criteria amongst the remaining WES data were then pooled to make the list of candidate variants for further filtering and study. The post processing of the raw sequence data and the filtration up to generating the list of co-segregating variants was conducted by our collaborator at the University of British Columbia, Canada.

2.3. Polymerase Chain Reaction Primer Design

Oligonucleotide primers were designed using sequence data obtained from the Ensembl Genome Browser database (http://www.ensembl.org). *Primer3* software version 4.0.0 (http://primer3.ut.ee) (Koressaar and Remm, 2007), and Primer-Basic Local Alignment Search Tool (BLAST) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to confirm the specificity of the primer binding sites, identify primer-primer complementarity and self-complementarity and calculate the primer melting temperatures. Forward and reverse PCR primer sequences, annealing temperatures and the expected sizes of the PCR products are recorded in **Tables 2.1 to 2.6**. The primers were synthesised by Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, SA.

The primers designed to validate the WES variants by Sanger sequencing and used to screen populations for the variant frequencies are recorded in **Table 2.1**.

Table 2.1 WES co-segregating variant primers for amplification and HRM

Chr	Gene	Primer	Sequence (5'-3')	T_m (°C)	T_a (°C)	Product size(bp)
1	***POU2F1***	Forward	AGTGAAGTTGGGTAGCTGGG	59,31	56,0	407
		Reverse	AGGCCCCACACTAACAATCA	58,93		
2	***CFAP65***	Forward	CCAAACTTGCCCTTCCTTGG	59,32	56,0	382
		Reverse	CTCCCACCTCAGCTTCTTGA	59,02		
3	***RFT1***	Forward	TTTTCCTGTCTCCTCCCCAG	58,64	55,6	368
		Reverse	GGGACAGGAGGAGGCAAAA	59,23		
11	***NRXN2***	Forward	TTTATCTCCGTGGTGCCCTC	59,46	56,2	388
		Reverse	AGCCAGGAGAGCTGTATGTG	59,17		
11	***ACTN3***	Forward	GCTGAGCTGAGACTTGAATCC	58,72	55,8	385
		Reverse	AAAGGTCATGAGAGGCAGCT	59,01		
14	***TEP1***	Forward	AGTCAGAGAAATGCAGGGGA	58,34	55,3	394
		Reverse	CCTTTCCCAGTGCTCTCAGA	59,02		
16	***CCNF***	Forward	AGTTCCCACGTGCTTCTCTT	59,24	56,0	154
		Reverse	AGCACATGAAGGGCAGGTGAC	58,87		
17	***CDC27***	Forward	GCAAACCGTTTCTCCATAATTCA	57,95	55,0	418
		Reverse	TGGAAAACAGGGAAGAAAGGG	58,04		
18	***TUBB6***	Forward	CCTTCATCGGCAACAGCAC	59,50	55,3	250
		Reverse	TCCGACTATCCATCGATCTCC	58,29		

Chr, chromosome; Tm, melting temperature; Ta, annealing temperature; *POU2F1*,POU class 2 homeobox 1; *CFAP65*,cilia and flagella associated protein 65; *CDC27*, Cell Division Cycle 27; *RFT1*, requiring fifty-three one homolog; *NRXN2a*, neurexin alpha 2; ACTN3, Alpha-actinin-3; CCNF, Cyclin F; TEP1, telomerase-associated protein 1; TUBB6, Tubulin Beta 6 Class V.

2.4. Polymerase Chain Reaction

The fragments of interest were amplified from either gDNA or complementary DNA (cDNA) in a 25µl PCR reaction containing 30 – 100 ng of template genomic DNA, 20 pmoles of both the forward and reverse primer, 75 µM dNTPs (Promega, Madison, Wisconsin, U.S.A), 1,5 mM MgCl2, 1X Green GoTaq® Reaction Buffer (Promega, Madison, Wisconsin, U.S.A) and 0,5 U GoTaq® G2 Flexi DNA Polymerase (Promega, Madison, Wisconsin, U.S.A). All PCR reactions performed included a non-template negative control to identify any PCR contamination. Amplification was performed in an ABI 2720 Thermal Cycler (Applied Biosystems Inc., Foster City, California, U.S.A). PCR conditions comprised of a denaturation step of 94°C, between 25 to 35 cycles of denaturation at 94°C, annealing at the primer's optimum temperature (**Table 2.1** to **Table 2.6**) followed by an extension step at 72°C. A final extension step of 72°C was performed for seven minutes before a cooling step at 4°C for 10 minutes. For lengthy fragments the extension time was dependent on amplicon length and complexity, generally an extension time of 60 seconds per 1000 bases was used.

2.5. Gel Electrophoresis

To visualise the PCR amplicons and to investigate if non-specific primer binding or contamination was present, gel electrophoresis was performed as follows: A 1% (w/v) agarose gel with 1µg/ml ethidium bromide (Sigma-Aldrich., United States) and 1X sodium borate (SB) buffer (was cast and immersed in 1X Sodium Tetraborate (SB) buffer (give buffer components i.e. final concentration). Multiple molecular size markers were used for the duration of the study to determine the size of the fragment size and 5 to 8 µl of PCR product was loaded into the gel wells. Electrophoresis, was performed at 10 V/cm for 25 to 30 min. To visualise the DNA fragment, the SynGene UV gel documentation system (Synoptics Ltd., Cambridge, UK) was utilised with GeneTools software version 3.0.6 (Synoptics Ltd., Cambridge, UK).

2.6. Post-Polymerase Chain Reaction Purification for Sanger Sequencing

Enzymatic PCR clean-up was done to remove the remaining primers and dNTPs from the reaction. Products were purified by mixing 8 µl of the PCR reaction, with one unit of Shrimp alkaline phosphatase (Sap I) (Cleveland, Ohio, USA) and 0,1 units of Exonuclease (Exo I) (Promega, USA). The reaction was then incubated in an ABI 2720 Thermal Cycler at 37 °C for 15 minutes and 80 °C for 15 minutes to inactivate the enzymes (Applied

Biosystems Inc., Foster City, California, U.S.A). The DNA concentration was subsequently determined using the NanoDrop™ 2000 (Thermo Scientific, MA, U.S.A) and the product was diluted to a concentration of 30 ng/µl.

Samples were sequenced at the Central Analytical Facility (CAF) (Stellenbosch University, Stellenbosch, South Africa). Capillary electrophoresis was done on a 3130 x1 Genetic Analyser (Applied Biosystems, CA, U.S.A) using The BigDye Terminator Sequence Ready Reaction kit version 3.1 (Applied Biosystems, CA, U.S.A) following in house optimised reaction conditions. Analysis of data was carried out using BioEdit v7.2.0 software (Hall, 1999). Sequencing data was aligned with reference sequences generated from Ensembl Genome Browser database (http://www.ensembl.org).

2.7. Bioinformatic *In-silico* Analysis of Variants

To investigate the genetic variants identified, comprehensive computational analysis was performed. Functional prediction was determined from sequence homology–based programs namely SIFT; (http://sift.jcvi.org/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (http://www.mutationtaster.org/) and CADD (http://cadd.gs.washington.edu/). To determine evolutionary constraint acting on genomic sites GERP++ was used (http://mendel.stanford.edu/SidowLab/downloads/gerp/).

SIFT and PolyPhen-2 both reports result in terms of pathogenic scores accompanied by the prediction, for readability only the predictions are reported. MutationTaster reports the amino acid change impact as one of two predictions of either neutral or pathogenic based on its disease-causing threshold. CADD reports phred-like scores (“scaled C-scores”) ranging from zero to 50.Variants predicted to be amongst the 10 %, 1 % or 0,1 % most deleterious substitutions that can occur in the human genome, at that specific base position, are assigned a score of between 10 and 19, 20 and 29 or 30 and 50 respectively. Thus, the higher the CADD score the more likely it is that the variant is highly pathogenic. GERP++ uses maximum likelihood evolutionary rate estimation for position-specific scoring, and the score ranges from -12,3 to 6,17. The closer the score is to 6,17(representing the most conserved a region can be), the greater the level of evolutionary constraint inferred to be acting on that site.

2.8. Population Frequency Evaluation

2.8.1. High Resolution Melt Analysis

High resolutions melt (HRM) analysis was used to screen ethnic matched controls for the variants identified. For detection a 1/100 dilution of SYTO® 9 (Thermo Fisher, Waltham, MA, U.S.A), a green fluorescent nucleic acid stain was added to the standard PCR protocol as described in section 2.4. HRM analysis detects the change in fluorescence as a double stranded PCR product dissociates to single-stranded DNA with increasing temperature. Initially, the fluorescence will be high, but reduces gradually as the double stranded DNA intercalating dye is released. The resulting thermal denaturation profile is unique to that specific PCR product, because DNA strand melting depends on sequence length, bases and GC content (HRM Assay Design and Analysis Booklet; http://www.corbettlifescience.com/shared/RotorGene%206000/hrm_corprotocol.pdf). This enables visual discrimination between homozygous and heterozygote genotypes (**Figure 2.1**). The melt temperature conditions are from 65°C to 95°C rising by 0,1°C increments on a Rotor-Gene 6000 analyser (Corbett Life Science, Australia).

In the population screening experiments, the proband sample with the known variants was included as a reference for comparison. For every variant, any sample exhibiting altered thermal denaturation profiles on HRM was selected for Sanger sequencing to characterise the sequence variation. In a validation step samples would be randomly selected from each variant run for sequencing. In total for each screened variant 20 control samples were sequenced (10,4 % of total controls). A non-template control was included in each run to monitor contamination.

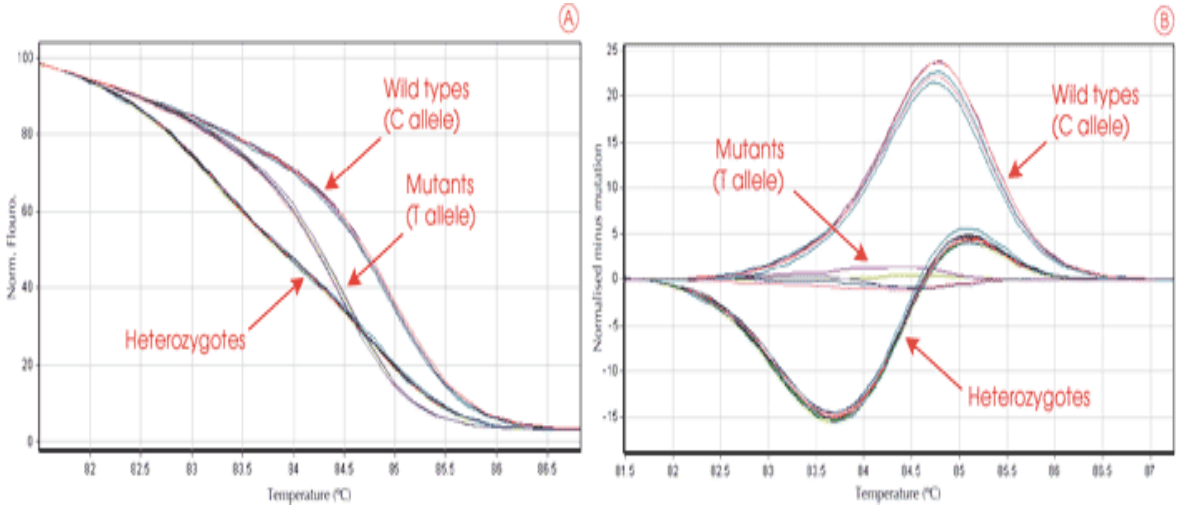


Figure 2.1 HRM curve analysis. It is possible to differentiate between variants in human ACTN3 (R577X) SNP genotypes using HRM. Homozygous wild-type (C allele), mutation (T allele) and heterozygote samples are shown on a standard normalised melt

curve (A) and a difference plot normalised to mutant samples (B). Amplification and HRM analysis were done using a Rotor-Gene 6000 instrument and genotypes were automatically assigned by the Rotor-Gene software (Adapted with public domain rights from http://hrm.gene-quantification.info).

2.8.2. Online Population Frequency Databases

The frequency of genetic variants in other populations was determined by searching public mutation databases. Each database comprises of reported sequencing data from multiple populations, the frequency at which alleles occur in any given population can then be determined. This is referred to as the population specific MAF, and it enables us to differentiate between common and rare variants in populations. Variants with a MAF greater than or equal to 0,01 are considered to be polymorphisms, being found at a frequency greater than 1 % in the general population. The population frequency databases examined included the ExAC database (http://exac.broadinstitute.org/), the 1000 Genomes Project (http://browser.1000genomes.org/index.html) and dbSNP (https://www.ncbi.nlm.nih.gov/snp). Populations examined were the European (Non-Finnish), African, East Asian, Latino and South Asian.

2.9. Pathway and Expression Analysis

To further prioritise the variants according to gene function and interactions, pathway analysis was performed using Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathways Analyser (http://www.genome.jp/kegg/pathway.html) and PANTHER Pathway Analyser (http://www.pantherdb.org/pathway/). The publicly available expression databases accessed were the Allen Brain Atlas (www.brain-map.org/) and Human Protein Atlas (https://www.proteinatlas.org/). Both databases record human mRNA expression data, but Allen Brain Atlas gives only brain regional whole-transcriptome gene expression data.

B. Investigating the Prioritised Gene

Although computational analysis predicted the pathogenicity of the candidate variant in NRXN2α, characterisation of the variant and its protein is necessary to validate the in-silico predictions.

2.10. Frequency of *NRXN2a* Variant in Patient Group

The HRM analysis was performed on the Rotor-Gene 6000 analyser (Corbett Life Science, Australia) to screen 671 SA PD patients for the NRXN2 c.G3008A (p.G849D) variant. The

same *NRXN2α* primers that were used to screen the controls were used for the patient screening. PCR reactions were prepared according to the standard PCR protocol as described in section 2.4. For detection 1 µl (1/100 dilution) of SYTO® 9 (Thermo Fisher, Waltham, MA, U.S.A), was added to reaction mix. The melt temperature conditions ranged from 65°C to 95°C rising by 0,1°C increments.

A proband sample with the known variants was included as a reference for comparison in each run to determine the samples that could be exhibiting altered thermal denaturation profiles. Thereafter, these sample would be sent for Sanger sequencing to characterise the observed sequence variation. As with the control screening, samples were randomly elected from each run for sequencing. In total 30 patient samples were Sanger sequenced (5 % of the total screened patients). A non-template control was included in each run to monitor contamination. Patient samples that were unsuccessfully amplified or that exhibited abnormal thermal denaturation profiles were repeated.

To screen the prioritised gene in PD patients from other populations Annex (https://annex.can.ubc.ca/about), a web-based private database (from our collaborator Prof. Farrer) was searched. The database comprises of over 3000 exomes of patients with different neurological diseases from all over the world. Access to the database was granted by the system administrator and a user account was provided. The data contained has been deidentified and each user can search the database but there is no option to download the exome data files. The database was searched for NRXN2α variants that were specifically found in diagnosed PD patients. The repository has an inbuilt frequency database called Genome Aggregation Database (GnomAD) (gnomad.broadinstitute.org/) and also reports CADD scores for each variant.

2.11. Protein Structure Homology Modelling

The initial step of the homology modelling was to identify templates of protein structures evolutionary related to the target sequence. This was accomplished using the SWISS-MODEL Webserver (https://swissmodel.expasy.org/) (Bordoli and Schwede, 2012) which identifies suitable templates through a BLASTP (Basic Local Alignment Search Tool) search. BLASTP (https://blast.ncbi.nlm.nih.gov/BLAST.cgi?PAGE=Proteins), is an algorithm used for computing primary biological sequence information, such as the amino acid sequences of different proteins (Altschul et al., 1997). The search enables a researcher to compare a query sequence with a library or database to identify sequences that resemble the query sequence above a certain threshold. The sequence search was performed in SWISS-MODEL using default parameters with the NRXN2α target

sequence (Uniprot ID number Q9P2S2, Transcript ENST00000265459.10) as the search query.

The selection of a template from the available files was based on properties like sequence identity, structural similarity, Global Model Quality Estimation (GMQE) score and the Qualitative Model Energy Analysis (QMEAN) score. A NRXN2α protein PDB (Protein Data Bank) structure file was found and selected based on its properties for further analysis. The model was evaluated against its respective template using standalone application PyMol (Schrödinger L, 2015) and the SWISS-MODEL inbuilt sequence to structure workbench SWISS-PDB Viewer (Guex and Peitsch, 1997) to ensure the quality of the model. These programs allow for the superimposing of two structures to find the differences in the backbone atoms, stereochemistry, spatial distribution of small charged groups and main chain hydrogen bonding (Moreland et al., 2005).

2.12. Prioritised Gene Primers

Initially a gene amplification primer pair (**Table 2.2 A**) was designed to bind to the 5' and 3' ends of *NRXN2α* (**Figure 2.2 A**). The primer set was designed on the basis of the following criteria: 50-60 % GC content, 25-35 nucleotides long and an annealing temperature below 60ºC. Due to the initial primer set not amplifying full-length *NRXN2α* from cDNA, shorter gene primers that would bind to the 5' and 3' ends of the gene were designed to improve the specificity of binding. Overlapping internal primers that were 20 nucleotides long, with melting temperatures of 56ºC and GC contents of 55 % were also designed to use with the shorter gene primers (**Table 2.2 B**). We expected that the two separate fragments amplified from the cDNA library could be ligated to form full-length NRXN2α cDNA. The primer binding sites are illustrated in **Figure 2.2 A** and **Figure B.**

Table 2.2 *NRXN2α* Gene amplification primers and PCR conditions

***A.* *NRXN2α* Gene Primers with Restriction sites**

Gene	Primer	Sequence (5'-3')	T_m (˚C)	T_a (˚C)	Product size(bp)
***NRXN2α* EcoR1 5'A**	Forward	ACTGCAGAAGAATTCATGGCGTCC GGGAG	54,06*	51,0	5139
***NRXN2α* SalI 3'A**	Reverse	ACTGCAGAAGTCGACTCAGACATA ATACTCCTTGTCTTTG	56,62*		

***B.* Shortened *NRXN2α* Gene Primers and internal primers**

Gene	Primer	Sequence (5'-3')	T_m (˚C)	T_a (˚C)	Product size(bp)

***NRXN2a* EcoR1 5'B**	Forward	CAGAAGAATTCATGGCGTCCGGGAG	54,06*		
***NRXN2a* Internal Primer Rev**	Reverse	ATGATGCCCGTCTCAATGTT	55,5	51,0	2594
***NRXN2a* Internal primer Fwd**	Forward	AGCTGTCTGTGGACAACGTG	60,53	53,6	2633
***NRXN2a* SalI 3'B**	Reverse	GACGTCGACTCAGACATAATACTCCTTGTCTTTG	56,62*		

Turquoise, universal seat sequence, purple, EcoRI restriction enzyme site; yellow, SalI restriction enzyme site, red, start codon; green, stop codon; T_a, annealing temperature, T_m, melting temperature; ˚C, Degrees Celsius, Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

In an effort to increase the reverse transcription (RT) efficiency and specifically target the *NRXN2a* gene in cDNA synthesis from the Human Foetal Brain RNA library, we decided to use *NRXN2a*-specific primers in the initial step of RT (**Table 2.3**). A combination of four primers were designed that would bind to the 3´untranslated region (UTR) of the *NRXN2a* RNA, and these would be used instead of the random hexamers that were supplied with the QuantiNova™ Reverse Transcription kit, for cDNA synthesis. The resulting cDNA would be used to generate the full-length gene using the primers in **Table 2.2**.

Table 2.3 *NRXN2a* specific 3' UTR reverse transcription primers

Gene	Primer	Sequence (5'-3')
***NRXN2a* 3' UTR 1 Rev**	Reverse	TGAGGCAGCCAGGGAGAG
***NRXN2a* 3' UTR 2 Rev**	Reverse	TTTTTGCGTTTCCTCTTCGT
***NRXN2a* 3' UTR 3 Rev**	Reverse	GTGGACGGCAGGAGGAAG

NRXN2a, neurexin alpha 2; UTR, Untranslated Region; Ta, annealing temperature, Tm, melting temperature; ˚C, Degrees Celsius, Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

As neither of the previously described primer pairs were able to amplify the full-length *NRXN2a,* intron flanking primers (**Table 2.4**) were designed to amplify short PCR products (size ± 500 base pairs) from the cDNA. This was to check if there was complete *NRXN2a* in cDNA generated from the gene specific RT. If a PCR product was produced it would mean that the synthesised cDNA library contained *NRXN2a* gene DNA fragments spanning multiple exons. The first primer set designed spanned from exon 2 to exon 7 and would produce a product of 475bp. The second primer set spanned from exon 19 to exon

22 and would produce a PCR product of 515bp. The primer binding sites are shown in **Figure 2.2 C.**

Table 2.4 *NRXN2a* intron flanking primers

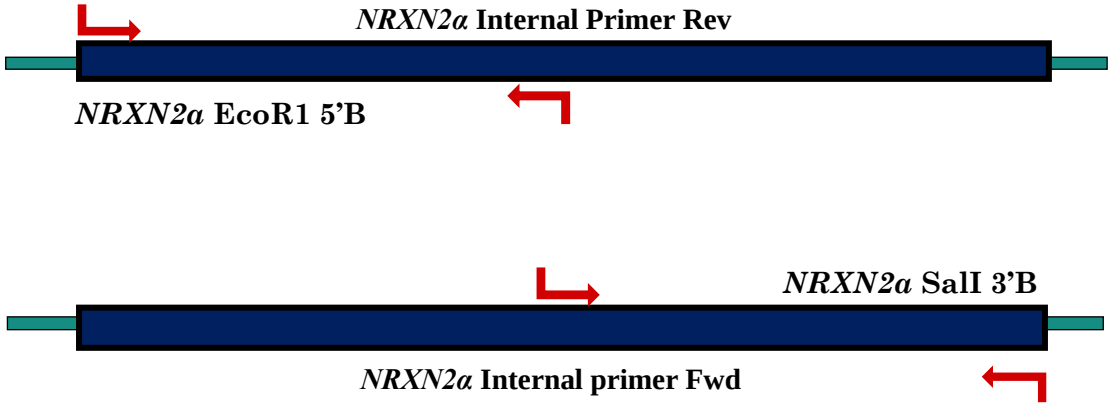
Gene	Primer	Sequence (5'-3')	T_m (˚C)	T_a (˚C)	Product size(bp)
***NRXN2a* Exon 2 For**	Forward	GCTTCGGCGGCAAGTTCT	61,05	55,9	475
***NRXN2a* Exon 7 Rev**	Reverse	ACCATAGCGTGTCCAATCCC	58,92		
***NRXN2a* Exon 19 For**	Forward	TGGTGCGCTTCACTCGAA	58,9	55,9	505
***NRXN2a* Exon 22 Rev**	Reverse	CCTCCAGGTCCTCATCATCG	59,0		

Tm, melting temperature; Ta, annealing temperature; *NRXN2a*, neurexin alpha 2; UTR, Untranslated Region; *, Tm calculated without restriction sites or enzyme seat; Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

Figure 2.2 is a schematic representation showing the *NRXN2a* gene with the multiple PCR amplification primers and their corresponding binding sites. This is to illustrate the different ways in which we attempted to amplify the full-length *NRXN2a* gene.

A. *NRXN2α* full length gene primers

B. Shortened *NRXN2α* gene primers with internal primers



C. *NRXN2α* intron flanking primers

Figure 2.2 *NRXN2α* primer binding sites schematic diagram. The relative locations of primer binding sites are indicated by the red arrows, the full-length gene is illustrated by the blue and the untranslated region is shown by the turquoise.

2.13. cDNA Synthesis

An expression vector containing the prioritised gene was required to perform functional studies. To generate this expression vector, full-length cDNA of the prioritised gene was synthesised using the QuantiNova™ Reverse Transcription kit (Qiagen, Hilden, Germany). Before cDNA synthesis, 1 µg total RNA from a commercially available Human Foetal Brain RNA library (CLO- 636526, Takara Bio Europe, Saint-Germain-en-Laye, France) was treated with 1x gDNA wipeout buffer and RNase free water on the 2720 Thermal Cycler (Applied Biosystems, CA, USA) according to the manufacturer's instructions, to ensure no DNA contamination. cDNA synthesis was carried out with the purified RNA following the manufacturer's instructions. The RT reaction was performed

using 2720 Thermal Cycler (Applied Biosystems, CA, U.S.A). A volume of 14 µl of purified RNA was added to 1 µl RT Primer Mix, 4 µl 5 × Quantiscript RT Buffer and Quantiscript reverse transcriptase (100 units/µl) before incubating at 42°C for 30 min. The RT reaction was terminated by incubating at 95°C for three min, the synthesised cDNA was then used as the template for PCR amplification. An internal control supplied with the RT kit was also included in the run as a positive control for synthesis.

2.14. cDNA Quantification and Quality Assessment by Absorbance

cDNA quantity and quality were determined by a NanoDrop spectrophotometer 2000 (Thermo Fisher Scientific, United States) using the DNA mode. The quantity of the sample is reported ng/ul, the quality is determined using the ratio of absorbance at 260nm and 280nm and 230nm. It is generally accepted that a 260/280 ratio that is greater than 1,80 indicates adequate purity and integrity of DNA. The cDNA was stored at -20°C until required for PCR amplification reactions.

2.15. cDNA Quality Assessment by Polymerase Chain Reaction

2.15.1. Amplification of Exon 1 To Exon 2 of HBB from cDNA

To further verify the purity of the cDNA and detect possible gDNA contamination, the Haemoglobin subunit beta (*HBB*) gene was amplified. The *HBB* s primer pair was previously designed to span a small intron (130 bp) between exon 1 and exon 2 (**Table 2.5 and Figure 2.3**). This PCR therefore allows discrimination between cDNA and gDNA on the basis of product size.

Table 2.5 Primer sequences for HBB reference gene

Name of Primer	**Primer Sequence (5'-3')**	**Tm (˚C)**	**%GC**	**Ta (˚C)**	**Size of PCR Product (bp)**
HBB-F	ACACAACTGTGTTCACTAGC	58	45	55	gDNA: 377 cDNA: 247
HBB-R	ACCGAGCACTTTCTTGCCAT	60	50		

F, Forward; R, Reverse; Ta, annealing temperature, Tm, melting temperature; ˚C, Degrees Celsius, Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

A. HBB Primers spanning intron

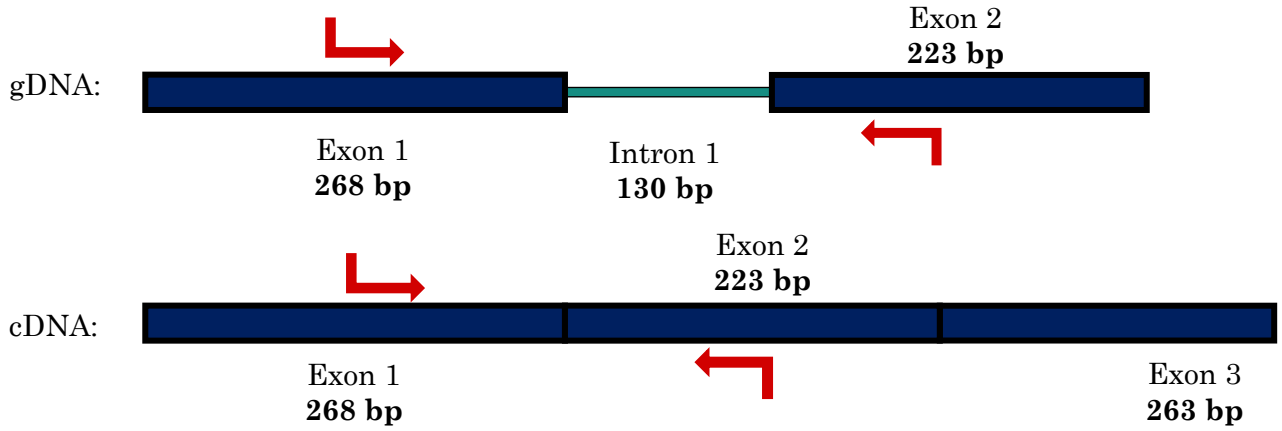


Figure 2.3 Schematic diagram of the amplification of *HBB* from cDNA and gDNA. Forward and reverse primers are highlighted in red. Intron 1 to 2 is 130 bp in size. After PCR the gDNA product will be 377 bp in size and the cDNA PCR product will be 247 bp because the intron is spliced out.

2.15.2. Amplification of Exon 2 To Exon 5 of PARK2 from cDNA

PARK2 is considered to be one of the largest genes in the human genome, this is due to its large introns (Asakawa et al., 2001). The primers which had been previously designed, amplify a region spanning from exon 2 to exon 5, flanking three introns with a combined total size of 389,151bp (**Figure 2.4**). The primer sequences are shown in **Table 2.6**. Due to the splicing of introns in RNA these primers are expected to produce a cDNA PCR product of 419bp whereas it there should be no PCR product from gDNA as amplifying the 389,151 bp product would not be feasible.

Table 2.6: Primer sequences for *PARK2*, exon 2 to exon 5.

Name of Primer	Primer Sequence (5'-3')	Tm (˚C)	%GC	Ta (˚C)	Size of PCR Product (bp)
***PARK2* Exon 2 For**	CACCATTTAAGGGCTTCGAG	60	50	55	gDNA: 389,151 cDNA: 419
PARK2 Exon 5 Rev	TTCCTGGCAAACAGTGAAGA	58	45		

For, Forward; Rev, Reverse; Ta, annealing temperature, Tm, melting temperature; ˚C, Degrees Celsius, Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

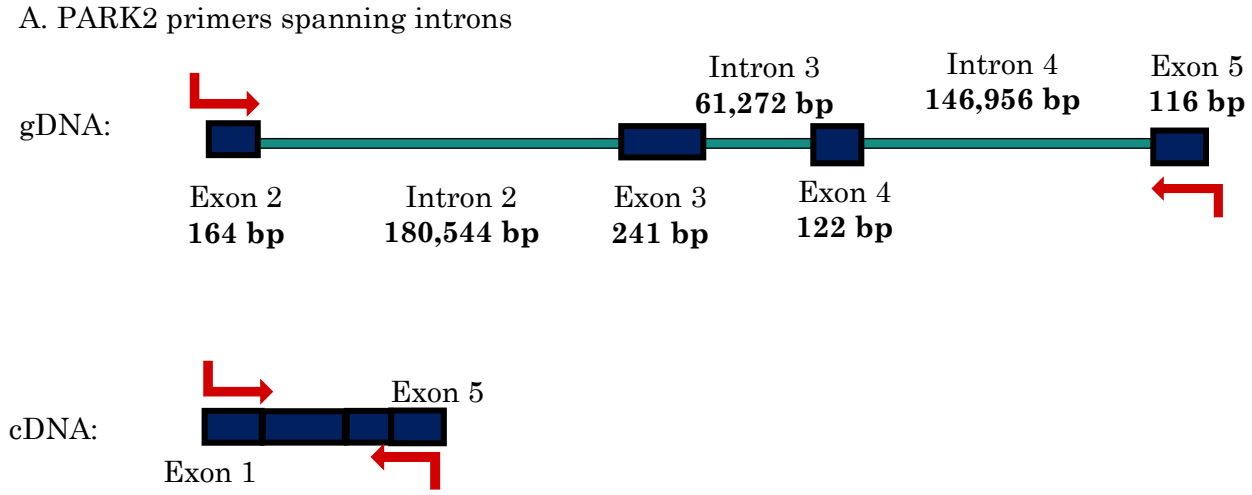


Figure 2.4 *PARK2* PCR cDNA and gDNA schematic diagram. Forward and reverse primers are in red. After PCR the cDNA product will be 419 bp in size and there should not be any PCR amplification in gDNA possible due to extremely large introns.

2.16. Culturing SH-SY5Y Cells for Protein Expression Studies

2.16.1. Thawing of SH-SY5Y Cells from Frozen Stocks

The human Neuroblastoma SH-SY5Y adherent/suspension cell line (Cellonex, Johannesburg, RSA) (passage 12) originally derived from metastatic bone marrow, were rapidly thawed from long-term storage of frozen aliquots at -80°C. The vial containing the frozen stock was immersed for three to five minutes in a 37°C water bath. Thereafter, the outside of the vial was immediately sterilized with 70% Ethanol and placed in the Esco Airstream® Class II Biohazard Safety cabinet (Esco Technologies Inc., United States).

2.16.2. Removing Freezing Media from Frozen Stocks

Culture media was prepared using Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F-12) in a ratio of 1:1 (Lonza, Walkersville, MD, USA), supplemented with 20% Foetal Bovine Serum (Lonza, Walkersville, MD, USA) and 1% of the antibiotic Penicillin-Streptomycin (Lonza, Walkersville, MD, USA). The antibiotic mixture is added to prevent contamination by gram-positive and gram-negative bacteria (Kohanski et al., 2010). The culture media was then prewarmed to 37°C and 5 ml of prewarmed culture media was added to the thawed cells in a 15 ml Eppendorf tube. Cells were spun at 1500 rpm for 5 minutes and the supernatant was discarded. The pelleted cells were resuspended in 5 ml of culture media and seeded in a 25 cm^2 flask (Corning Inc., United States) with an additional 10 ml of culture media. Thereafter, the flask was gently swirled to distribute the cells evenly over the growth surface of the flask and incubated at 37°C in a 5% carbon dioxide humidified incubator (Farma International, Miami, Florida, U.S.A).

For growth and maintenance of the cells, the culture media was changed when the cells became adherent and cell debris was evident.

2.16.3. Sub-culturing of SH-SY5Y Cells

The cells were sub-cultured when 80 to 90% confluency was reached, and this typically took 3 to 4 days. As a majority of the SH-SY5Y cells were adherent to the growth surface, the culture media was removed from the flask and the cells were gently rinsed with 2 to 3ml of sterile Phosphate Buffered Saline (PBS) (Lonza, Walkersville, MD, U.S.A), at room temperature. The cells were treated with 3ml Trypsin (0,5g/L Trypsin) (Lonza, Walkersville, MD, U.S.A) for 5 minutes to facilitate the detachment of cells from the growth surface of the flask. Culture media (5ml) was then added to the cell suspension, this was then transferred to a 15 ml Eppendorf tube. The cells were spun at 1500 rpm for 5 minutes and the supernatant was discarded. The pelleted cells were resuspended in 6 ml of culture media and 3ml of the resuspended cells was seeded into each 25cm^2 flask (Corning Inc., United States) with an additional 12 ml of prewarmed culture media.

2.16.4. Lysis of SH-SY5Y Cells

To prepare whole cell protein extracts from the SH-SY5Y cells, 1X Cell Lysis Buffer (CLB) was prepared using 1M Tris- HCl at pH 7,4, 5M NaCl, 1% of the total volume of Triton™ X-100 (Sigma-Aldrich, United States), 1M MgCl2 and ddH_2O to a final volume of 1L. The cOmplete™ EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, United States) was added immediately before using the CLB. To harvest the cells the culture media was removed, and the flasks were washed once with 4ml ice-cold PBS. A volume of 1,5 ml ice-cold PBS was added to each flask, and a cell scraper was used to scrape the cells off the growth surface. The detached cell suspension was transferred to 2ml Eppendorf® LoBind microcentrifuge tubes (Eppendorf, United States). The tubes were centrifuged at 15,000 rpm for 5 minutes, and the supernatant was gently removed with a pipette. The pellet was resuspended in 200 µl ice-cold CLB containing the Protease Inhibitor Cocktail and ~0,5 ml of sterile 0,1 mm glass beads before incubating on ice for 30 minutes. Subsequently the lysates were homogenized at 4°C using a Precellys ® 24 tissue homogenizer (Bertin Instruments, Amsterdam, Netherlands) three times for 30 seconds with 5 minutes intervals on ice. The lysates were the centrifuged at 15,000rpm at 4°C in a bench top centrifuge (Labnet International Inc., United States) for 30 minutes. The supernatant was carefully transferred to a new 2 ml LoBind microcentrifuge tube and stored at -80°C.

2.16.5. Bradford Assay for Determination of Protein Concentration in Cell Lysates

The protein concentration of cell lysates was determined using the Bradford assay to enable the loading of equal amounts of protein per well for western blot (WB) experiments. Bovine Serum Albumin (BSA) (Biolabs, New England) standards ranging from 0-1000 ng/ml were prepared and 10µl of each BSA standard was pipetted in duplicate in a microtiter plate. Volumes of 1/20, 1/10 and 1/3 of lysate samples were then pipetted in duplicate and a lysis buffer control was added. Next, 200µl Bradford reagent was added to the BSA standards and lysate samples with a five-minute incubation step before the plate was read at 595nm absorbance on a Synergy HT luminometer (BioTek Instruments Inc., Vermont, U.S.A) using the KC4™ v3.4 software (BioTek Instruments Inc., Vermont, U.S.A). The BSA standard curve was generated and the average lysate concentration was determined from the technical duplicates for each sample.

2.17. Western Blot Analysis of Endogenous NRXN2α Expression

2.17.1. Membrane Blocking and Antibody Binding

After running the precast sodium dodecyl sulphate (SDS) gel in a polyacrylamide gel electrophoresis (PAGE) system, the proteins were transferred to nitrocellulose membrane. The automated IBLOT™ transfer system (Invitrogen™, United States) was used, with iBlot™ Transfer Stacks containing integrated nitrocellulose transfer membranes. Transfer of protein was carried out using the default transfer parameters of 7 minutes at 20-25 Volts. Once completed the membranes were removed and washed in Tris-buffered saline with Tween 20 (TBS-T) buffer (20 mM Tris at pH 7,4, 150 mM NaCl, 0,1% Tween 20) for five minutes. The membrane was then incubated for an hour in 5% BSA to ensure the blockage of all non-specific binding sites. The membrane was rinsed in TBS-T, cut appropriately and incubated overnight at 4°C in 5 ml of the appropriate antibody diluted in TBS-T on a Stuart® orbital shaker SSL1 (Barloworld Scientific Ltd., United Kingdom).

Following overnight incubation, the membrane was rinsed four times with TBS-T for a total wash time of 45 minutes, while shaking at room temperature on a rotating shaker (Labcon Pty, Ltd, Maraisburg, RSA). Next, the membrane was incubated for 1 hour at room temperature with the relevant horseradish peroxidase (HRP) conjugated secondary antibody. Thereafter, the membrane was again washed three times in TBS-T for a total wash time of 30 minutes.

2.17.2. Chemiluminescent Detection

Detection of the membranes was done in the dark room. The reagents of the Clarity™ Western Enhanced Chemi-Luminescence Blotting Substrate (Bio-Rad Laboratories, United States) were mixed as follows; 5 ml of Clarity western luminol/enhancer reagent and 5 ml Clarity western peroxide reagent to make a 1:1 working solution. The antibody treated membranes were incubated in the solution for 5 minutes. The membranes were then placed in an autoradiography cassette, covered with a transparent plastic sheet avoiding bubbles. A CL-Xposure™ autoradiography film (Thermo Fisher Scientific, United States) was placed in the cassette and the exposure time varied between 10 seconds and 10 minutes, depending on the strength of the signal. The exposed films were developed using a Hyperprocessor™ automatic autoradiography film processor (Amersham Pharmacia Biotech Ltd., United Kingdom) and the protein bands were visualised.

2.17.3. Optimisation of Primary Antibody Dilution

The Anti-Neurexin II alpha (NRXN2α) antibody (rabbit, polyclonal), Anti-Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody (Goat, Polyclonal) and HRP-conjugated secondary antibodies were all obtained from Abcam (BIOCOM Biotech, South Africa). The optimal dilution of each of the antibodies was experimentally determined. For the Anti-NRXN2α antibody 1/400, 1/500 and 1/1000 dilutions were tested **(Table 2.7)**,while for the Anti-GAPDH antibody 1/1000 and 1/3000 dilutions were tested.

Table 2.7 Characteristics of the relevant antibodies used in WB analysis

Target Protein	Antibody	Recommended Primary Ab dilution	Optimised Primary Ab dilution	Secondary Ab dilution	Expected size (kDa)
NRXN2α (ab34245)	Rabbit/polyclonal	1:400	1:500	1:5000	185
GAPDH* (ab9483)	Goat/polyclonal	1:1000	1:3000	1:5000	37

Abbreviations: Ab: antibody, kDa –kilodalton, *represent the loading control

2.17.4. Quantification of Western Blot Signal

All Western blots were quantified using the open source platform, ImageJ (http://imagej.nih.gov/ij/) developed by the National Institute of Health. The exposed western blotting films were digitalized by scanning and saved as .jpg files. Before quantification the jpg images need to be reduced from a 24-bit to an 8-bit images using the jPOR custom palette provided by ImageJ. Each band is converted to a peak in which the intensity of the signal is proportional to the peak area. The protein band intensity can

now be measured by highlighting the area of each peak and drawing a baseline across the bottom of the peaks, thereby excluding the underlying area from the measurement. ImageJ employs the pixel intensity of the bands and to quantify the relative abundance of the protein on the WB. All blots were performed in biological triplicates and GAPDH was used as the internal control protein. This was necessary to verify that the samples were loaded uniformly on the gel and to enable normalisation of the relative protein concentrations amongst the samples.

Chapter Three
3. Results
A. Whole Exome Sequencing Results and Prioritisation

In this section we use WES to screen a multiplex family with PD, then validate the list of variants using Sanger sequencing. Computational analysis is then performed on the validated variants to predict the pathogenicity before prioritising the list of candidates to a single gene.

3.1. Identifying the Family

This study used a WES approach to screen a multiplex family with PD for candidate PD-causing genes. Identification of the family and the expansion of the pedigree was done by the consulting neurologist, Prof. Jonathan Carr (Tygerberg Academic Hospital, Cape Town, SA). The full pedigree of this family, which is denoted as family ZA253, contains 57 individuals (**Appendix II, Figure 1**). In total, 11 members were reported to have been affected by PD. A simplified version of the pedigree indicating the family members who took part in this study, is shown in **Figure 3.1.**

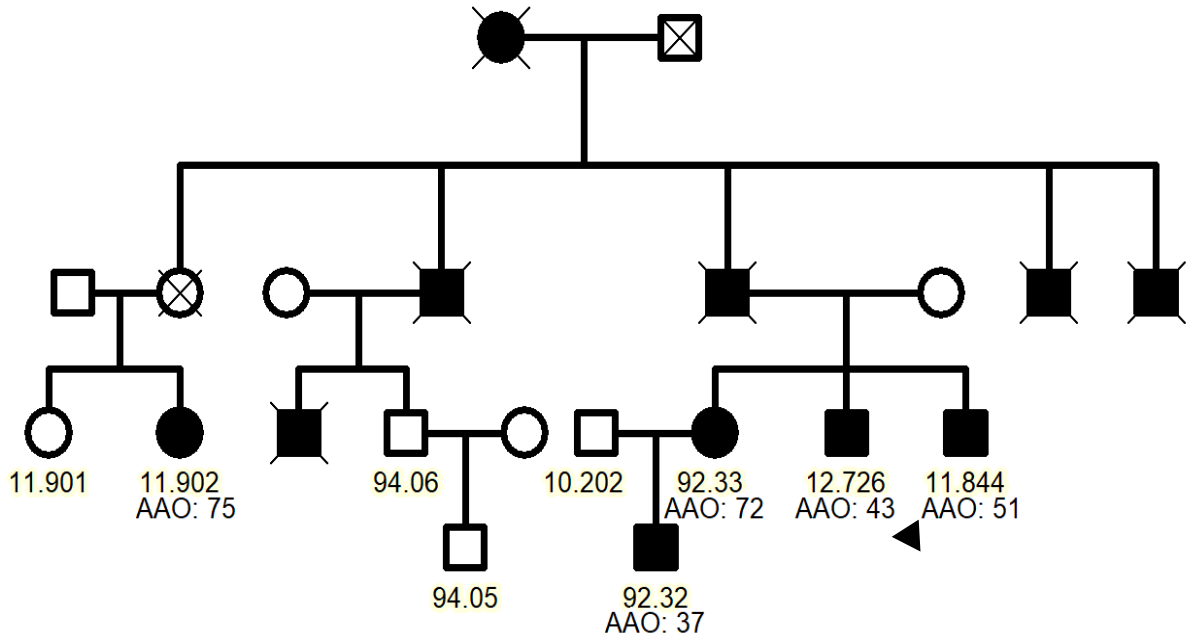


Figure 3.1 Pedigree of family ZA253. Circles denote females and squares depict males. The filled in symbols indicate affected individuals. The intersecting diagonal lines indicate that the person is deceased. The numbers below each individual is the laboratory ID number. For readability and confidentiality, the pedigree is greatly simplified. Branches without medically confirmed PD or without DNA samples were omitted, but all known living family members with PD diagnoses are included in the pedigree. AAO, Age at onset, the black arrow shows the proband.

Patient 11.844 (Proband) was the first individual to be formally diagnosed with PD at the age of 51 years. He presented with typical PD with some of the classical motor symptoms including tremor, rigidity and balance problems. Shortly after the diagnosis, his siblings were assessed, and individual 12.726 was also diagnosed with PD, and his age of onset is estimated to be 43 years. The typical PD motor symptoms were noted without any NMS, and the cognitive assessment showed no cognitive decline. Patient 92.32, the nephew of the two affected siblings was later examined by Prof. Carr and diagnosed with early-onset PD at the age of 37 years. His most prominent motor symptom was early dystonia. His mother (individual 92.33) was examined by Prof. Carr and diagnosed with PD with an AAO of 72 years.

Patient 11.902 was diagnosed with PD in 2017, Prof. Carr noted that this individual developed a slowly progressive form of the disease. She had an AAO of 75 years and the predominant symptoms at the time of diagnosis were tremor and REM sleep behaviour disorder. A summary of the clinical characteristics of each family member is recorded in **Table 3.1.**

Table 3.1 Clinical information of multiplex Afrikaner SA Family

	Affected				*Unaffected*				
Individual ID	11.844* (proband)	12.726* (sibling)	92.32* (nephew)	92.33 (sibling)	11.902^ (cousin)	94.06 (cousin)	94.05 (nephew)	11.901 (cousin)	10.202 (spouse of sibling)
Gender	M	M	M	F	F	M	M	F	M
Age in 2018	63	69	44	72	80	69	43	71	73
Disease Diagnosis	PD	PD	PD	PD	PD*	Unaffected	Unaffected	Unaffected	Unaffected
AAO	51	43	37	72	78	NR	NR	NR	NR
Clinical Signs	Typical PD	Slow Duration dystonia, minimal autonomic and cognitive decline	Early Dystonia	Late-onset Mild PD	Tremor, Rapid eye movement sleep behaviour disorder	NR	NR	NR	NR

M, male; F, female; AAO, age at onset; NR, not relevant; PD, Parkinson's disease; ^; denotes potential phenocopy of disease; *; individuals WES was conducted on.

3.2. Whole Exome Sequencing

WES was performed on three of the affected family members (12.726, 11.844 and 92.32). The WES metrics, reported in **Table 3.2**, revealed good overall coverage of the sequencing in all three individuals. Individuals 92.33 and 11.902 had not yet been diagnosed with PD at the time that WES was performed. For the data analysis, we made the following assumptions that the three affected individuals had the same genetic cause of disease and the mode of inheritance was AD. Importantly, none of the known causal mutations associated with PD were found in any of the affected individuals. This indicates that it is plausible that there is a novel PD-causing mutation in this family.

Table 3.2 Summary of metrics of WES in the three affected individuals

Metrics	WES Individuals		
	12.726	**11.844**	**92.32**
Total number of reads	136,048,746	127,298,302	114,750,902
Non-duplicated reads	90,255,052	97,195,484	85,612,771
Reads aligned to target	82,045,320 (90,9%)	83,570,035 (86,0%)	74,665,623 (87,2%)
Mean target coverage (%)	88,12	79,88	71,04
Total number of variants	22,318	22,566	22,531

All synonymous co-segregating variants were excluded. This left a total of nine novel or rare non-synonymous variants that were shared between the three affected individuals (**Table 3.3**). The genes in no particular order are *NRXN2a* (Neurexin 2 alpha), *ACTN3* (Actinin Alpha 3), *POU2F1* (POU Class 1 Homeobox 1), *CFAP65* (Cilia and Flagella Associated Protein 65), *RFT1* (Requiring Fifty-Three 1), *TEP1* (Telomerase-Associated Protein 1), *CCNF* (Cyclin F), *CDC27* (Cell Division Cycle 27), and *TUBB6 (Tubulin Beta 6 Class V)* .

Table 3.3 Rare and novel exonic variants shared between individuals 12.726, 11.844 and 92.32 after variant filtering.

Chr	Gene	Genomic coordinate	Ref	Alt	Exon	Variant Type	Amino Acid position	dbSNP Accession ID
1	*POU2F1*	167343435	G	C	6	Non synonymous	p.A154P	rs774001310
2	*CFAP65*	219886565	A	G	18	Non synonymous	p.T1023A	rs748008106
3	*RFT1*	53126455	C	G	12	Non synonymous	p.A463G	novel
11	*NRXN2α*	64418979	G	A	13	Non synonymous	p.G849D	novel
11	*ACTN3*	66319116	A	C	2	Non synonymous	p.E127A	rs569022261
14	*TEP1*	20871567	A	G	7	Non synonymous	p.Y412C	novel
16	*CCNF*	2495617	G	C	10	Non synonymous	p.C363S	novel
17	*CDC27*	45249391	A	C	3	Non synonymous	p.Y48S	rs62077270
18	*TUBB6*	12326073	A	C	4	Non synonymous	p.T429P	rs745469423

Chr, chromosome; p., protein codon position; rs, reference SNP ID number, Nucleobases - A, adenine; G, guanine; C, cytosine; Amino acids – A, Alanine; C, Cysteine; D, Aspartic acid; E, Glutamic acid; G, Glycine; P, Proline; S, Serine; T, Threonine; Y, Tyrosine. The genomic coordinates are from the GRCh37/hg19 reference genome, Nucleobases - A, adenine; G, guanine; C, cytosine; T, tyrosine.

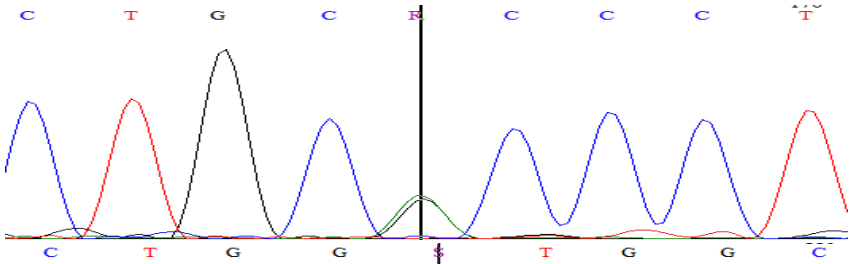
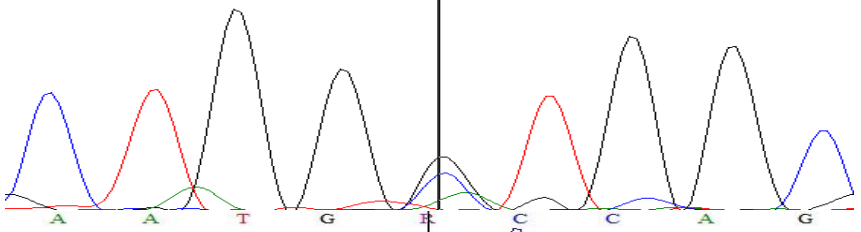
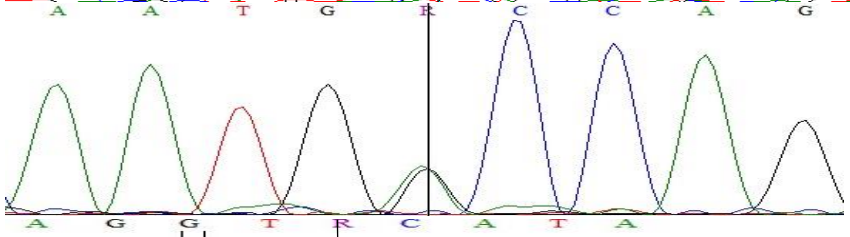
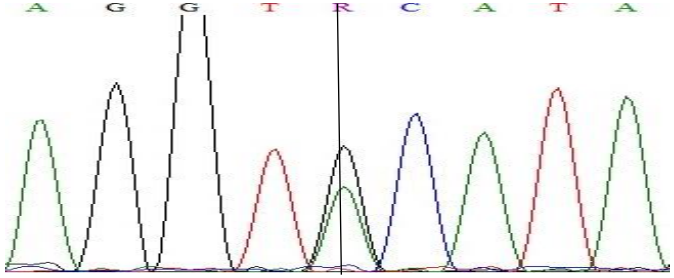
NRXN2a is thought to be involved in neuronal cell adhesion and in calcium channel activity regulation (Futai and Hayashi, 2008; Reissner et al., 2008). *ACTN3* functions as a structural component of sarcomere Z line regulating calcium ion binding and actin binding (Quinlan et al., 2010). *CDC27* forms a part of an essential ubiquitin-protein ligase called the anaphase-promoting complex without which the cell division cycle in mammalian cells stops at Metaphase (Topper et al., 2002).

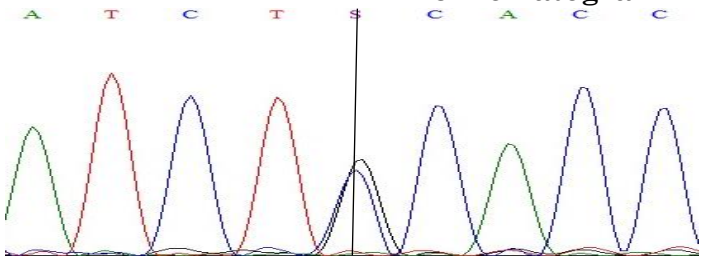
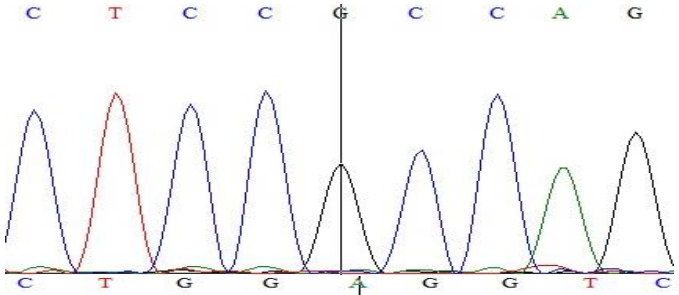
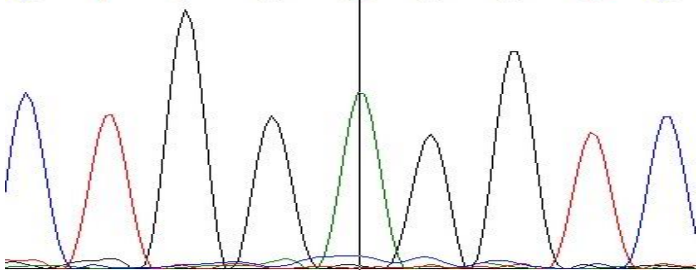
POU2F1 mainly regulates transcription of specific genes in the nucleus as a ubiquitous protein (Ng M. C. Y. et al., 2010). *TEP1* has been previously found to be a component of the telomerase complex, maintaining chromosome termini and shown to interact specifically with telomerase reverse transcriptase (Harrington et al., 1997; Weinrich et al., 1997). *CFAP65* is reported to be involved in regulating spermatogenesis and flagellar development (Tang et al., 2017).

RFT1 has been shown to encode an enzyme that catalyses the translocation of intermediate substrates from the cytosolic to the luminal side of the endoplasmic reticulum membrane (Haeuptle et al., 2008; Helenius et al., 2002). *CCNF* forms a part of a F-box motif which is an E3 ubiquitin-protein ligase complex mediating the ubiquitination of cellular substrates (Williams et al., 2016). The last of the nine genes *TUBB6* encodes a Tubulin polymer, which forms protofilaments, the major constituent of microtubules in the cytoskeleton (Sun et al., 2007).

3.3. Sanger Sequencing Validation of Variants

To ensure that the nine variants found using WES in the three affected individuals are not sequencing artefacts, Sanger sequencing was performed. The two newly diagnosed individuals and the unaffected family members whose samples were available, were also Sanger sequenced for co-segregation analysis. PCR was used to amplify the region around each variant from gDNA, and Sanger sequencing was performed in both the forward and reverse directions. The Sanger sequencing chromatograms generated from the screening of the nine variants in individual 11.844 (Proband) are shown in **Figure 3.2**. The variant genotyping results of the entire family are summarised in **Table 3.4**. The four WES variants that were found to be sequencing artefacts in the three affected individuals were subsequently excluded from the co-segregation analysis.

Gene	Sequence around variant	Chromatogram
		Validated Variants
CFAP65	CAACTGC[A]CCCTCT	
RFT1	CTGG[C]TGGCC	
NRXN2α	AATG[G]CCAG	
TEP1	AGGT[A]CATA	

Gene	Sequence around variant	Chromatogram
CCNF	ATCT[G]CACC	
		Non-Validated Variants
POU2F1	CTCC[G]CCAG	
ACTN3	CTGG[A]GGTC	

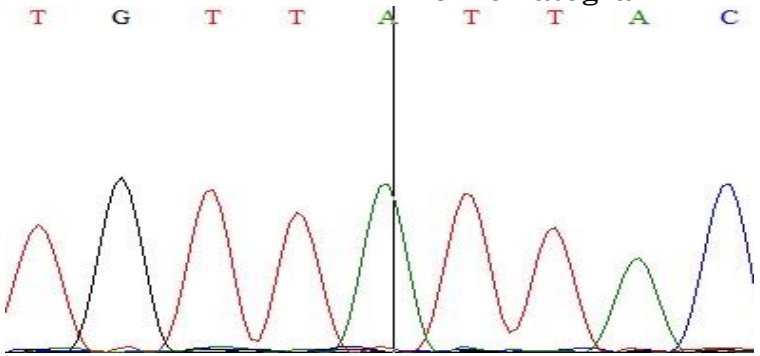
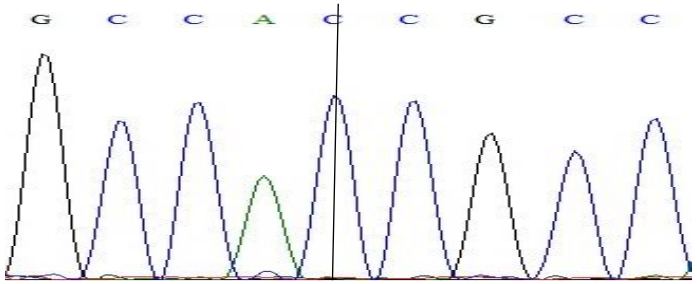
Gene	Sequence around variant	Chromatogram
CDC27	TGTT[A]TTAC	
TUBB6	GCCA[C]CGCCA	

Figure 3.2 Sanger sequencing validation of WES co-segregating variants. Chr, chromosome; Nucleobases - A, Adenine; G, Guanine; C, Cytosine, T, Tyrosine . Variants of interest are indicated by [ ] in the sequence column, and by a solid line through the variant base on the chromatograph.

Table 3.4 Genotyping of the five validated variants in the family.

		Gene				
		CFAP65	***RFT1***	***NRXN2a***	***TEP1***	***CCNF***
Affected individuals	11.844 (proband)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
	12.726 (sibling of proband)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
	92.32 (nephew)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
	92.33 (sibling)	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
	11.902 (cousin)	**Yes**	No	No	No	No
Unaffected individuals	94.06 (cousin)	No	No	No	No	**Yes**
	94.05 (nephew)	No	No	No	No	**Yes**
	11.901 (cousin)	No	No	No	No	**Yes**
	10.202 (spouse of sibling)	No	No	No	No	No

The relationships that are shown for each family member are in reference to the proband; **Yes,** Variant is Present; No, Variant is Absent.

Of the nine prioritised variants, five (*CFAP65, RFT1, NRXN2a, TEP1* and *CCNF*) were validated by Sanger sequencing and four (*POU2F1, ACTN3, CDC27* and *TUBB6*) were considered sequencing artefacts. The five validated variants were all present in four of the affected individuals, one affected individual (11.902) had only one of the five validated sequence variants, in the *CFAP65* gene. The unaffected family members only shared one co-segregating variant with the affected individuals, in the *CCNF* gene, which indicates that this variant might be a rare polymorphism.

3.4. Bioinformatic *In-silico* Analysis Using Functional Prediction Tools

The amino acid changes for each variant are recorded in **Table 3.5**. None of the base changes reported occurred in the interchangeable third base position (wobble base) and therefore all resulted in non-synonymous exonic changes in the proteins. The *RFT1* A463G variant causes a transition from Alanine, a small non-polar, aliphatic amino acid to glycine which is also a small non-essential, non-polar amino acid. *TEP1* and *CCNF* both had the substitution from one polar group to another. The *TEP1* Y412C variant changes tyrosine, an aromatic amino acid with a polar side chain to Cysteine which is a smaller, non-charged, polar amino acid. The variant occurring in *CCNF* C363S changes Cysteine, a polar, non-charged amino acid to Serine which is a much smaller, polar amino acid with a hydroxymethyl side chain. The exchange of a polar group for a non-polar one was observed in both *CFAP65* and *NRXN2a*. The *CFAP65* T1023A variant causes Threonine which is a polar, uncharged amino acid to be substituted by an Alanine, a smaller non-polar, aliphatic amino acid. The *NRXN2a* G849D amino acid change was from glycine a small, non-polar, side chain free amino acid, to Aspartic acid a larger, negatively charged amino acid.

Various functional prediction tools were used to further prioritise the sequence variants. The results of the computational analysis performed are summarised in **Table 3.5** with the functional scores and predictions from SIFT, PolyPhen-2, MutationTaster, GERP++ and CADD indicated. The *TEP1* variant was predicted to be benign across all the different prediction tools with a CADD score of zero and with a gene GERP++ score of - 8,29. The conservation score indicates that in the region of the genome where this variant occurs, more substitutions occur than for the average neutral site, thus indicating that this region may not be under evolutionary constraint.

The variant occurring in *RFT1* was also predicted to be benign by all four tools, although though its conservation score implies that it occurs in a region under evolutionary constraint. The *CFAP65* and *CCNF* variants both had three out of the four different computational tools predicting them to be pathogenic, and the gene conservation scores were positive indicating that they both occurred in regions of the genome under evolutionary constraint. The only variant that was predicted to be pathogenic across all four functional prediction tools was *NRXN2a*, which had a CADD score of 29,50 and a conservation score of 4,93 showing that the level of evolutionary constraint inferred to be acting on this site is rather high.

Table 3.5 Computational tool scores and predictions for validated variants.

Chr	Gene	cDNA Pos	Amino Acid Pos	Codon Change	Gene Pos	Exon	SIFT Pred	PolyPhen2 Pred	Mut Taster Pred	GERP	CADD
2	***CFAP65***	c.A3151G	p.T1023A	**A**CC⇒**G**CC	exonic	18	Deleterious	Benign	Disease-causing	2,38	18,70
3	***RFT1***	c.C1450G	p.A463G	G**C**T⇒G**G**T	exonic	12	Tolerated	Benign	Neutral	2,53	4,96
11	***NRXN2a***	c.G3008A	p.G849D	G**G**C⇒G**A**C	exonic	13	Deleterious	Probably Damaging	Disease-causing	4,93	29,50
14	***TEP1***	c.A1276G	p.Y412C	T**A**C⇒T**G**C	exonic	7	Tolerated	Benign	Neutral	-8,29	0,00
16	***CCNF***	c.G1176C	p.C363S	T**G**C⇒T**C**C	exonic	10	Tolerated	Probably Damaging	Disease-causing	5,43	27,30

Chr, chromosome; Pos, position, p., protein codon position, Nucleobases - A, adenine; G, guanine; C, cytosine; Amino acids – A, Alanine; C, Cysteine; D, Aspartic acid; E, Glutamic acid; G, Glycine; P, Proline; S, Serine; T, Threonine; Y, Tyrosine.

3.5. Frequency of Variants in Controls

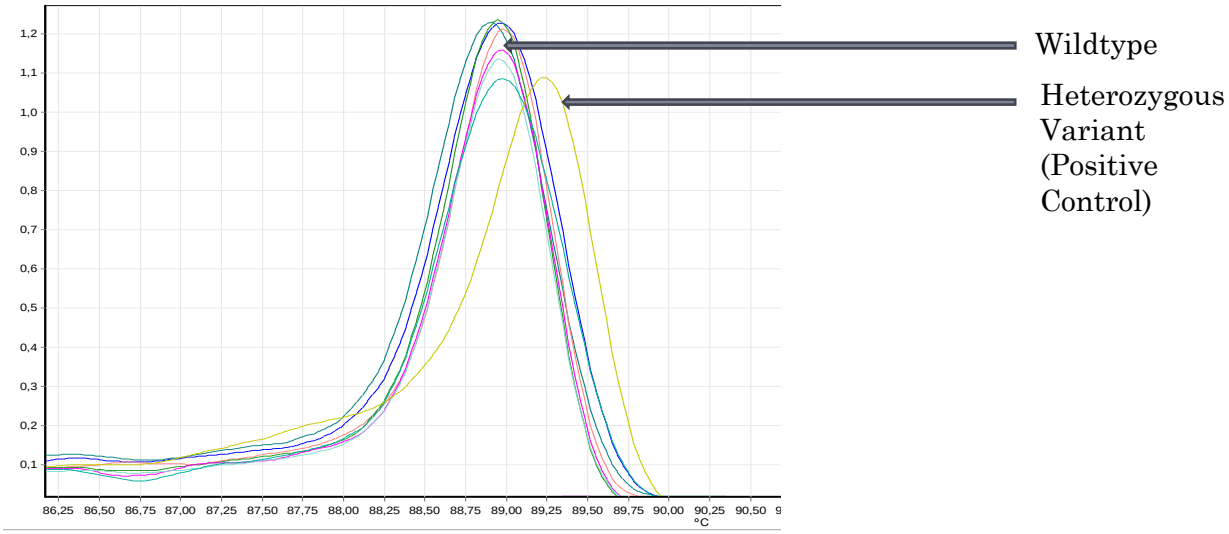
3.5.1. Screening Online Frequency Databases

To determine the frequency of the validated variants in multiple control populations an extensive search was performed in publicly available databases including the ExAC database and the 1000 Genomes Project (**Table 3.6**). Results from the databases revealed that four of the five validated variants had not been previously reported. This could mean they are either extremely rare or novel variants in the screened African American, American, East Asian, South Asian, Finnish and Non-Finnish European population exome data. The *CFAP65* variant had been previously reported and was found to be present in the ExAC Aggregated Population at a MAF frequency of 17 out of 2 100 000 individuals. In assessing the allele frequency of these variants, it is important to consider not just the publicly available online databases, but also to screen the appropriate ethnically-matched control individuals.

3.5.2. Screening the Local Population

To determine the allele frequency in ethnically-matched control samples, HRM analysis was performed on 192 SA Caucasian controls, 96 of which self-identified as Afrikaner. The HRM results for the *NRXN2a* variant are shown in **Figure 3.3**. For this variant, the reference allele is a guanine and the variant allele is an adenine residue. Both the melt curve **(Figure 3.3 A)** and the difference graph **(Figure 3.3 B)** distinguish the variant in the positive control (individual 11.844) from the wild-type allele in healthy controls**.** None of the 192 ethnically-matched controls had any of the five variants (**Appendix; Table 3).** The absence of the validated variants in the local control population means that the variants are rare.

A.



B.

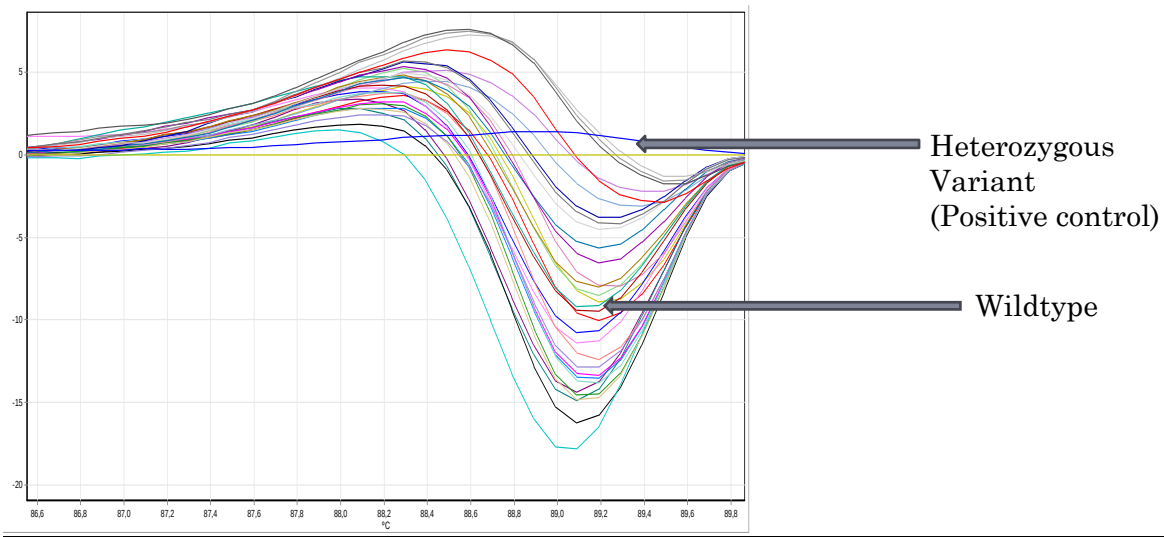


Figure 3.3 HRM analysis performed on the G849D variant in *NRXN2a*, showing that the positive control sample melts differently from all the wild-type controls. **A**, Melt curve showing the that the different genetic sequences melt at slightly different rates. **B**, The Difference graph shows the normalised fluorescence of all the samples subtracted from a selected reference sample at different temperatures. The positive control was selected as the reference in this difference graph. Fluorescence is plotted on the y-axis and temperature in degrees Celsius is plotted on the x-axis.

Table 3.6 Allele frequencies for the five variants in multiple control populations.

			Online databases										Present Study
Chr	**Gene**	**Amino acid Pos**	**1000 Gen AFR AMR**	**1000 Gen AMR**	**1000 Gen EAS**	**1000 Gen EUR**	**ExAC AFR AMR**	**ExAC AMR**	**ExAC EAS**	**ExAC Finnish**	**ExAC Non-Finnish EUR**	**ExAC AGG**	**HRM Results**
2	*CFAP65*	p.T1023A	0	0	0	0	0	0	0	0	0	34/4 200000	0/384
3	*RFT1*	p.A463G	0	0	0	0	0	0	0	0	0	0	0/384
11	*NRXN2α*	p.G849D	0	0	0	0	0	0	0	0	0	0	0/384
14	*TEP1*	p.Y412C	0	0	0	0	0	0	0	0	0	0	0/384
16	*CCNF*	p.C363S	0	0	0	0	0	0	0	0	0	0	0/384

Chr, chromosome; ExAC, Exome Aggregation Consortium; AFR, African; AFR AMR, African American, AGG, Aggregated population, AMR, American; EAS, East Asian; EUR, European; SA, South African. Amino acids – A, Alanine; C, Cysteine; D, Aspartic acid; E, Glutamic acid; G, Glycine; P, Proline; S, Serine; T, Threonine; Y, Tyrosine.

3.6. Pathway and Expression Analysis

To add further information about the possible causal role of the variants, gene expression profiles across the human body and various human brain regions were assessed using publicly available databases, namely the Allen Brain Atlas (Sunkin et al., 2013) and the Human Protein Atlas (Pontén F. et al., 2011). Pathway analysis was also performed using KEGG Pathways Analysis (Kanehisa and Goto, 2000) and PANTHER Pathway Analysis (Mi and Thomas, 2009) to determine if any of the genes containing the five of variants are co-expressed, co-regulated or co- localize with each other or any of the known PD genes. This analysis also shows if the genes containing the variants and any of the known PD genes are functionally related or impact a pathway of interest such as neuronal development, regulation and functioning.

The results in **Table 3.7** show that none of the five variants interacted with any of the known PD genes. NRXN2α was noted by both the Allen Brain Atlas and the Human Protein Atlas to be highly expressed in the brain and selectively in the substantia nigra, a key brain region in PD pathogenesis. Pathway analysis showed that NRXN2α was involved in pathways that regulate synaptic functioning, neurotransmitter secretion and neuronal cell to cell adhesion. RFT1 was the only other gene to be reported as highly expressed in the brain, it was implicated in pathways of protein metabolism and of the endoplasmic reticulum membrane network.

CFAP65 expression was highest in the ciliac processes in the testis, fallopian tube and lungs, and it was associated with motile cilia pathway. TEP1 is expressed in all tissues of the body with expression reported to be low in the brain. It is reported to be involved in apoptotic pathways and in assembling telomerase components in cell signalling. Lastly CCNF, which was had low expression in the brain, was highly expressed in immunological tissues such as the lymph nodes, skin, appendix and bone marrow. It was associated with pathways of ubiquitination and cell cycle regulation.

Table 3.7 Summary of the expression and pathway analysis for the five proteins

Protein	Amino Acid Pos	Allen Brain Atlas	Human Protein Atlas	KEGG Pathways Analysis	PANTHER Pathway Analysis
CFAP65	p.T1023A	○ NR	○ Testis ○ Fallopian Tube ○ Lung	○ Nonhomologous End Joining ○ Cyano-Amino Acid Metabolism	○ Motile Cilium
RFT1	p.A463G	○ Expressed in all brain cells	○ Cardiac Muscle ○ Skeletal Muscle ○ Cerebral cortex ○ Hippocampus ○ Caudate ○ Cerebellum ○ Placenta	○ Endoplasmic Reticulum Membrane Network ○ N-Glycan Biosynthesis	○ Metabolism of proteins ○ Endoplasmic Reticulum Membrane Network
NRXN2α	p.G849D	○ Paravermis ○ Substantia Nigra (Pars Reticulata)	○ Exclusively expressed in the brain ○ Highest in the Cerebral Cortex	○ Cell adhesion molecules ○ Tight junction ○ Synaptic regulation	○ Neuron cell-cell adhesion ○ Chemical synaptic transmission ○ Neurotransmitter secretion ○ Regulation of molecular function ○ Synapse assembly ○ Postsynaptic membrane assembly
TEP1	p.Y412C	○ Paravermis ○ Left, Lateral Hemisphere	○ Expressed in all tissues	○ Apoptotic Pathways	○ CCKR signalling map ○ p53 pathway

			○ Highest in Parathyroid gland and Duodenum	○ Telomerase Components in Cell Signalling	○ Inflammation mediated Cytokine Signalling Pathway
CCNF	p.C363S	○ Cingulate Gyrus ○ Left, Lateral Hemisphere	○ Tonsil ○ Lymph node ○ Skin ○ Appendix ○ Bone marrow ○ Cerebral cortex	○ Ubiquitin mediated proteolysis ○ Cell cycle	○ Ubiquitination and Proteasome degradation ○ Cell cycle regulation ○ Regulation of phosphate metabolic process

Chr, chromosome; Pos, position; KEGG, Kyoto Encyclopaedia of Genes and Genomes, NR, not reported

3.7. Selection of Functional Candidate from Gene List

Of all five prioritised genes, *NRXN2α* is the strongest candidate based on the following evidence. It was shared by the four affected individuals with a similar disease phenotype. All four of the sequence-based computational tools correlated in their prediction of this variant being a pathogenic mutation with CADD classifying it within 0,1% of the most possible pathogenic substitutions that could have occurred at that specific locus (**Table 3.5**). The genome region where this variant occurs was reported to be highly conserved by GERP++ and it had the most severe codon change from a non-polar amino acid to a polar one. The *NRXN2α* variant was absent in the unaffected family members (**Table 3.4**), the ethnicity-matched matched neurologically normal controls and the populations screened on the online databases (**Table 3.6**). NRXN2α protein was reported by the pathway and expression databases to function in neuronal cell-cell adhesion, neurotransmitter secretion regulation and synaptic regulation (**Table 3.7**) and found to be highly expressed in the Pars Reticulata of the substantia nigra, a key brain region in PD pathogenesis. This makes it an ideal candidate for further analysis.

Based on meeting the requirements of the prioritisation criteria in all aspects, *NRXN2α* was the top candidate gene to be considered for further study. Although *in-silico* tools have enabled efficient, informative and reliable filtration strategies to filter candidate genes for disease, it is vital to functionally validate the prioritised novel candidate gene to validate the biological significance of the genes in a relevant disease model.

B. Results of Prioritised Gene Studies

Although computational analysis predicted the pathogenicity of the candidate variant in *NRXN2α*, characterisation of the variant and its protein is necessary to validate the *in-silico* predictions.

3.8. Frequency of *NRXN2α* Variant in South African Parkinson's Disease Patients

3.8.1. South African Parkinson's Disease Patients

To determine if the *NRXN2α* c.G3008A (p.G849D) variant is present in unrelated PD patients, HRM analysis was performed on 671 SA PD patients of various ethnicities (Afrikaner Caucasian 192, non-Afrikaner Caucasian 96, Black 192, Mixed Ancestry 96 and Indian 75). As previously shown by the HRM screening of this variant in ethnically-

matched controls, the variant allele (A) can clearly be distinguished from the wild-type allele (G) in the melt curve and the difference graph (**Figure 3.3 A and B**).

As with the control screening experiments, the positive control was the patient, individual 11.844, in whom the variant was validated by Sanger sequencing. The HRM screening showed that the variant was not found in any of the 671 patient samples. For validation, one sample was selected from each run (30 samples in total) to be Sanger sequenced. The Sanger results confirmed the HRM analysis results, that none of the PD patients had the *NRXN2a* c.G3008A variant allele.

3.8.2. International Cohort of Patients with Parkinson's Disease

To further screen a different subset of PD patients for other variants in the *NRXN2a*, Annex (https://annex.can.ubc.ca/about) was searched. The search for *NRXN2a* variants that were specifically found in diagnosed PD patients on this database yielded 24 genetic variants that were shared amongst 50 patients (**Table 3.8**). The only individuals who had the p.G849D variant in the database, are the affected individuals from our family who were sequenced by Prof. Farrer. All 24 variants are exonic and a total of 12 variants of the 24 were indels of various lengths. Majority of the variants were novel and not present on the inbuilt frequency database GnomAD with most CADD scores above 20.

Table 3.8 The screening results of exome data base Annex for *NRXN2a* variants

Genome Position	Ref/Alt	Variant	Zygosity	No. of PD patients	Unaffected (Non-PD)	dbSNP	CADD	GnomAD
64374723	G/T	p.A649D	Het	1	0	Novel	12,6	NA
64375161	AC/GA	c.1507_1508TC	Het	1	0	Novel	NA	NA
64375171	C/CCG	c.1498_1498delinsCGG	Het/Hom	3	0	Novel	NA	NA
64375212	G/T	p.P486H	Het	1	0	Novel	9,56	NA
64375283	C/CA	c.1386_1386delinsTG	Hom	1	0	Novel	NA	NA
64375321	C/G	p.A450P	Het	1	0	Novel	21,7	NA
64375390	/AG	c.1279_1279delinsCT	Hom	2	0	Novel	NA	NA
64375473	G/C	p.P399R	Het	4	0	Novel	13,9	NA
64387795	T/A	p.E362V	Het	1	0	Novel	28,3	NA
64390345	C/A	c.915G>T:p.E305D	Het	2	1	rs370812331	20,9	1,6e-5
64390450	A/C	c.810T>G:p.N270K	Het	1	0	Novel	25,1	NA
64393933	A/C	c.3847+2T>G	Hom	1	0	Novel	24,4	NA
64415814	G/C	c.3284-4C>G	Het	1	0	rs555038789	10,3	4,4e-4
64416245	C/CA	c.3124_3124delinsTG	Het	1	0	Novel	NA	NA
64416356	T/C	p.S1005G	Het	7	4	rs148473653	20,2	1,8e-3
64418793	G/C	p.P911R	Het	2	0	Novel	25,9	NA
64418979	**C/T**	**p.G849D**	**Het**	**3**	**0**	**Novel**	**29,3**	**NA**
64428522	C/T	p.G599S	Het	2	0	rs561880923	33,0	9,0e-5

Genome Position	Ref/Alt	Variant	Zygosity	No. of PD patients	Unaffected (Non-PD)	dbSNP	CADD	GnomAD
64434997	C/T	p.R477H	Het	5	0	rs140588352	24,4	8,0e-4
64480499	CG/C	c.672_673G	Het	1	0	Novel	NA	NA
64480574	C/T	c.598G>A:p.A200T	Het	4	0	rs797045803	19,7	7,2e-4
64480610	G/C	c.562C>G:p.P188A	Het	3	8	Novel	0,758	NA
64480970	AG/GC	c.201_202GC	Het	1	0	Novel	NA	NA
64480978	AGC/A	c.192_194T	Het	1	0	Novel	NA	0,0e+0

3.9. Structural Analysis of NRXN2α Protein with and Without Mutation

In the protein structural analysis, we used *in-silico* modelling to predict the effect of the NRXN2α G849D variant on protein structure. The amino acid change occurs in the laminin G-like domain four. The variant residue Aspartic acid is much larger than the wild-type residue glycine. Also, the glycine residue is a non-polar, side chain free amino acid while the variant introduces a negatively charged amino acid with a carboxyl group side chain (CH2COOH) (**Figure 3.4**). The introduction of Aspartic acid could lead to unwanted interactions with the neighbouring residues and could possibly affect the function and activity of the protein.

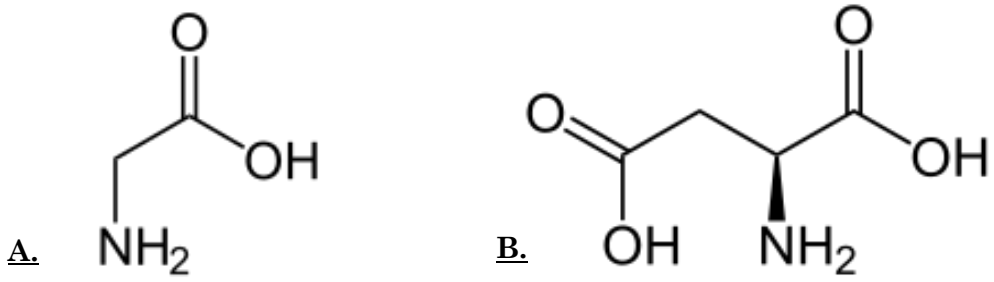


Figure 3.4 NRXN2α G849D amino acid residues. The amino acid structures of A. Glycine (G) and B. Aspartic acid (D).

To model NRXN2α, the relevant template was identified on Swiss-Model (https://swissmodel.expasy.org/) using the Uniprot ID number Q9P2S2 (Transcript ENST00000265459.10) for the longest human NRXN2α protein. The homologous template PDBID 3qcw (Neurexin 1 alpha, from *Bos Taurus*) was chosen for structure prediction because it shared 71,60 % sequence identity and over 52% structural similarity with the target protein. The full crystal structure of the NRXN2α protein is still to be solved so the NRXN1α template was chosen instead.

The GMQE score of the model was 0,62 (closer to one indicates greater reliability) and the QMEAN score was -0,86 (the closer to zero the more accurate the model) which suggests that the predicted model was reliable and accurate **(Figure 3.5)**. The replacement residue did not appear to affect the secondary structure of the protein (α-helix), however the amino acids do greatly differ in size .

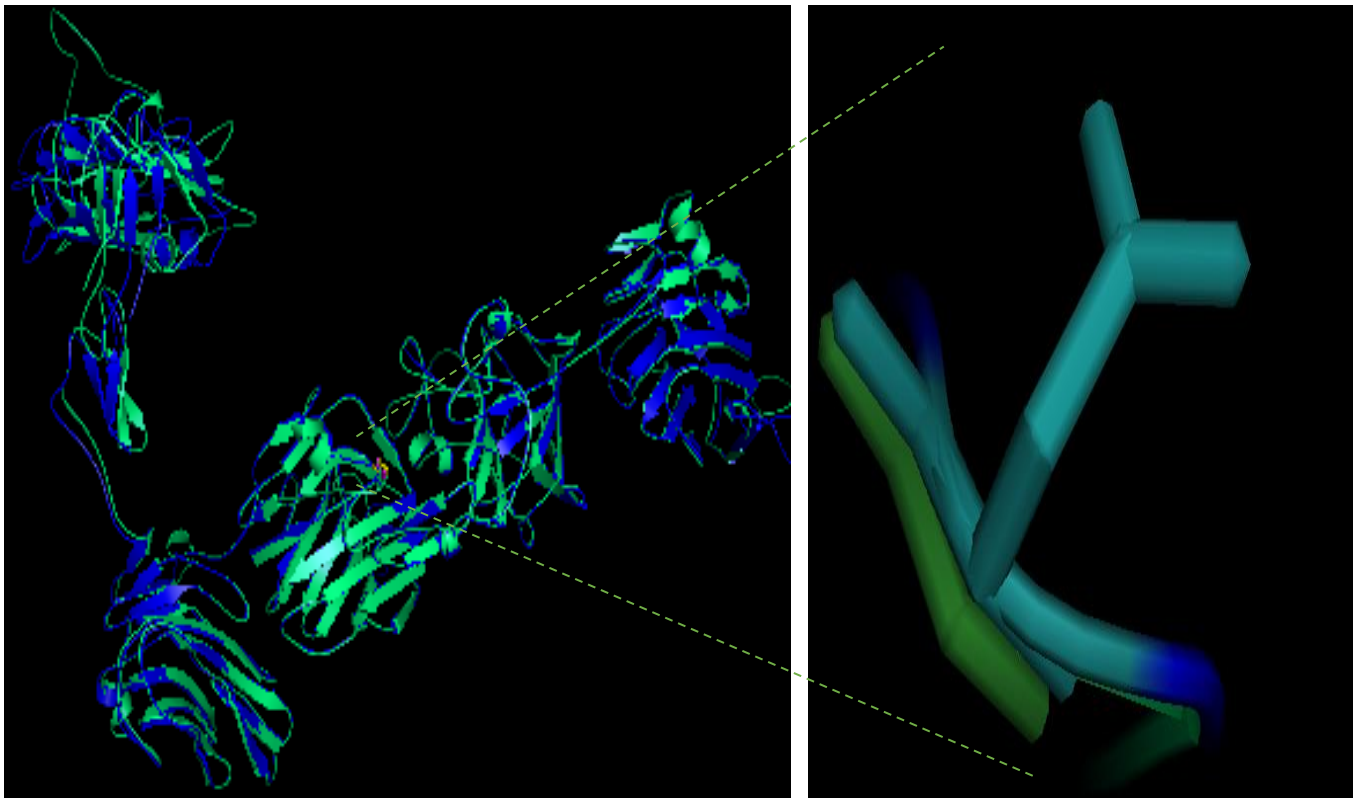


Figure 3.5 Superimposition of the predicted NRXN2α G849D model onto the homologous template structure the wild-type residue, Glycine (G), is shown in green and the mutant residue, Aspartic acid (D), is shown in light blue.

3.10. Generation of cDNA from Foetal Brain RNA Library

Full-length NRXN2α cDNA is not commercially available. The approach in this study was therefore to generate full-length NRXN2α cDNA using reverse transcription of the Human Foetal Brain RNA library. The variant form of the protein would then be generated by site directed mutagenesis to facilitate functional studies, in which the variant impact versus the wild-type could be assessed in a neuronal cell line.

3.10.1. cDNA Quality Assessment by Spectrophotometry and Gel Electrophoresis

To generate a full-length gene product of *NRXN2α*, cDNA was synthesised using the QuantiNova™ Reverse Transcription kit (Qiagen, Hilden, Germany) from a commercially available Human Foetal Brain RNA library. The kit contains a RT Primer Mix that consists of a unique blend of oligo-dT (specificity for poly(A) tails) and random primers (can potentially anneal to any complementary RNA in the sample). This ensures that RT reaction will yield both full-length cDNA from RNA with poly(A) tail and fragmented cDNA templates from the library (including those generated from degraded RNA).

The concentration and purity of the cDNA generated, was measured by a NanoDrop spectrophotometer. An internal RNA control supplied with the RT kit was also included in the analysis to verify that the experiment worked. The measurements were taken using the 'DNA mode' and a no template control that was included in the RT was used as a blank. The concentration of the cDNA and the positive control was 2939,7 ng/ul and the RT internal control was 2971,1 ng/ul **(Table 3.9).** This suggests that the RT reaction worked well. The cDNA had a 260/280 ratio value of 1,85. This is the ratio of absorbance at 260nm and 280nm which is conventionally used to assess the purity of the nucleic acids, and it is generally accepted that a 260/280 ratio that is greater than 1,80 indicates adequate purity and integrity of DNA. The 260/230 ratio is a second measure of DNA purity, with the value of pure nucleic acids being within the range of 1,8-2,2.

Table 3.9 cDNA concentration and purity results obtained from NanoDrop

Sample	Nucleic acid concentration	Unit	260/230	260/280
cDNA Library	2939,7	ng/ul	1,78	1,85
RT internal control	2971,1	ng/ul	1,77	1,84

The NanoDrop can give an indication of the purity of nucleic acids however it cannot distinguish between cDNA and gDNA contamination. Therefore, agarose gel electrophoresis was performed to distinguish between gDNA and cDNA on the basis of fragment size (**Figure 3.6**). The cDNA sample was loaded at two different amounts and a gDNA sample was included as a reference. Since cDNA was synthesised using RT random primers and oligo-dT primers, it is expected that there will be cDNA fragments of varying lengths, which will appear as a smear. When gDNA is not subjected to any sort of PCR amplification or shearing it remains very large in size, due to size exclusion it would not be able to move into the gel. This would result in a single solid band in the well where it was loaded.

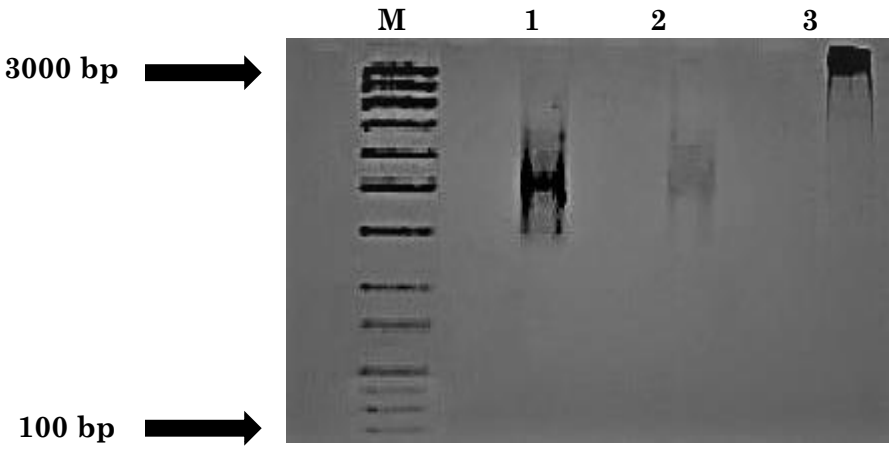


Figure 3.6 Agarose gel electrophoresis to distinguish between genomic DNA and complementary DNA. cDNA generated from a Human Foetal Brain RNA library was run at two different concentrations on a 1% agarose gel. In lane 1 contains 800 ng of cDNA, lane 2 contains 400 ng cDNA at and lane 3 contains 90 ng of gDNA. M is a 100 bp DNA ladder (O'RangeRuler™ DNA Ladder, Thermo Scientific).

Figure 3.6 shows that there was no gDNA contamination in the cDNA. Both the cDNA lanes appear to be smears, but lane 2 was faint. The smear shows the different sized fragments that were synthesised from the RNA library.

3.10.2. cDNA Quality Assessment by Polymerase Chain Reaction

3.10.2.1. Amplification of Exon 1 To Exon 2 of HBB from cDNA

To further verify the purity of the generated cDNA from possible gDNA contamination, a section of the Human Beta-globin (*HBB*) gene was amplified. The *HBB* primer pair had been previously designed to span a small intron (130 bp) between exon 1 and exon 2. The presence of this small intron in gDNA will generate a product of 377 base pairs (bp), while the cDNA, which lacks the intron, will generate a product of 247 bp (**Figure 2.3**). As expected, the cDNA template yielded a PCR product of between 200 to 300 bp in size and the gDNA produced a PCR product of between 300 to 400 bp (**Figure 3.7**).

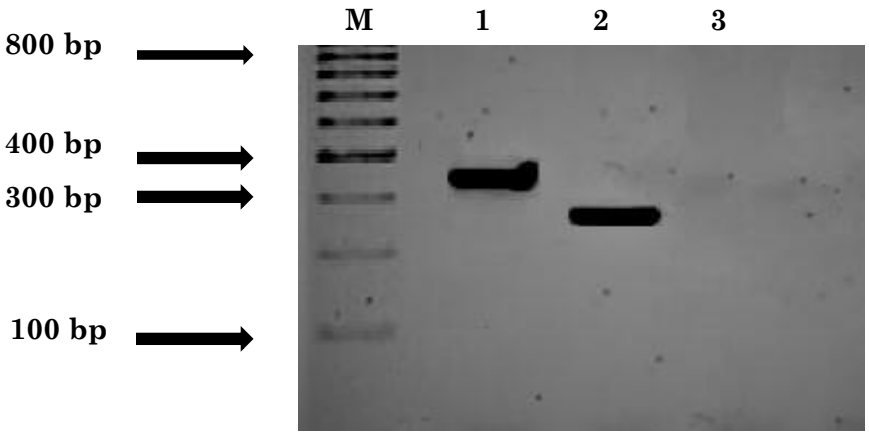


Figure 3.7 Visualisation of HBB products on a 1% agarose gel. In lane 1 is the gDNA product between 300-400bp, in lane 2 is the cDNA product between 200-300bp and in lane 3 is the no template control (NTC). A 100 bp ladder was used (O'RangeRuler™ DNA Ladder, Thermo Scientific).

The HBB amplification yielded products of the expected size indicating that there was no gDNA contamination in the synthesised cDNA. Sanger sequencing of the PCR products was conducted to confirm that the fragments obtained corresponded to the correct fragments. The sequencing confirmed that the observed products were indeed from *HBB*, and the sequence aligned to the reference at the correct size (**Figure 3.9**). The successful amplification from the cDNA also demonstrates that the RT experiment did successfully generate cDNA, as Taq polymerase are not able to amplify from RNA.

These quality assessments are important because gDNA contamination might result in false-positive results in downstream analysis or increase the chances of non-specific binding of the primers in subsequent experiments. Since the *HBB* primers span only one intron, it was decided to use another primer set that spans multiple exons in order to further evaluate the quality of the cDNA.

3.10.2.2. Amplification of Exon 2 To Exon 5 of PARK2 from cDNA

To assess if there was RNA degradation in the Human Foetal Brain Library, the amplification of a longer product that spans multiple exons was necessary. Primers for another gene, *PARK2* had been previously designed to amplify a region spanning from exon 2 to exon 5. For cDNA, these primers would produce a PCR product of 419bp whereas for gDNA the introns would result in a product 389 151 bp product which would not amplify by PCR (**Figure 2.4)**. **Figures 3.8** shows that the 419bp PCR product was successfully amplified from the cDNA. Sanger sequencing confirmed that the sequence corresponded to the correct region in *PARK2* gene (**Figure 3.10**). Whereas the gDNA template did not produce a band as expected.

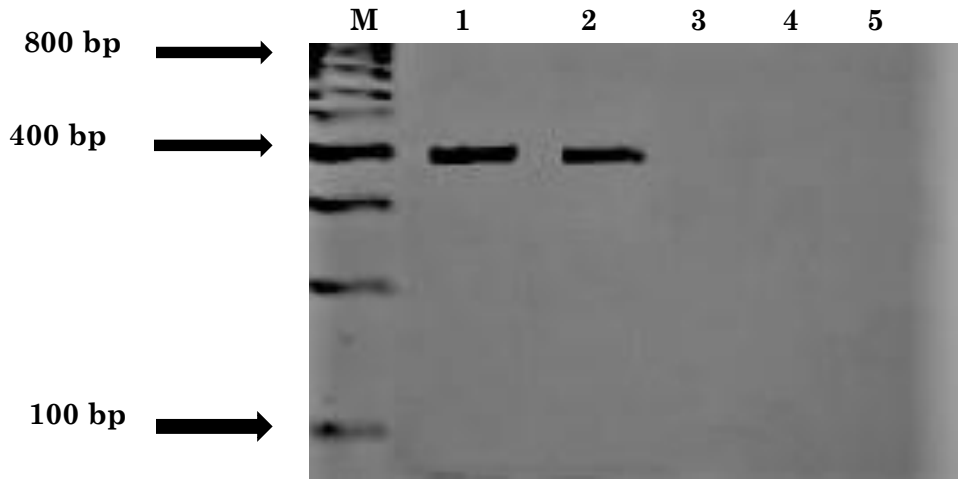


Figure 3.8 Visualisation of *PARK2* PCR amplification products on a 1% agarose gel. In lane 1 and 2 are the PCR products from the synthesised cDNA, in lane 3 and 4 are the gDNA products and the NTC is in lane 5. M is a 100bp ladder (O'RangeRuler™ DNA Ladder, Thermo Scientific).

These experiments verified that the synthesised cDNA was not contaminated with gDNA, that the RT of the Foetal Brain Library did synthesise cDNA fragments and these fragments spanned multiple exons. Taken together, this means that in theory, the cDNA is of sufficient quantity and quality to generate full-length *NRXN2α* cDNA.

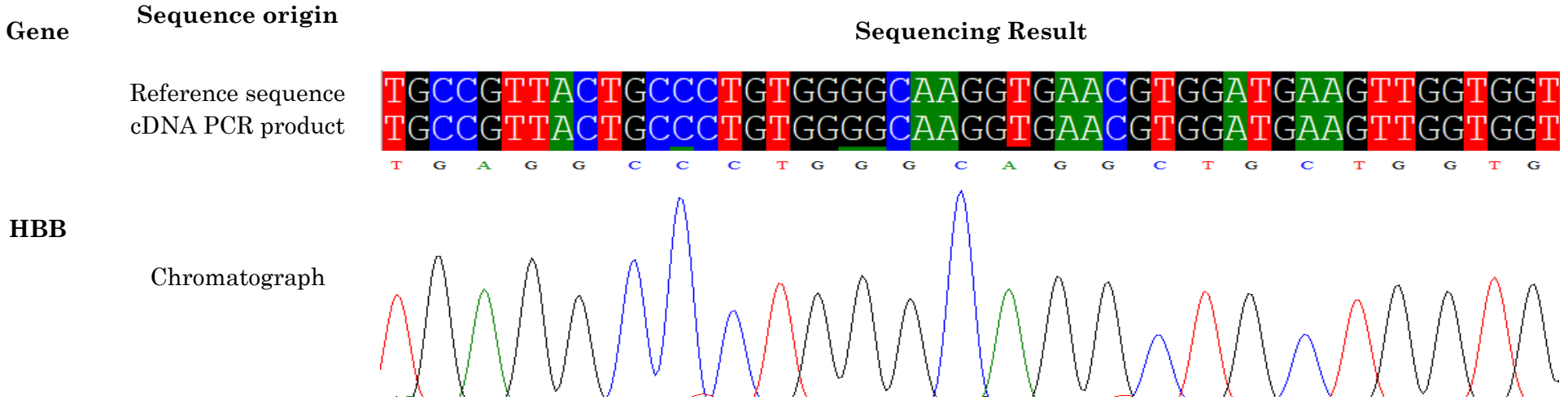


Figure 3.9 Partial sequence chromatograph and alignment of the *HBB* cDNA fragments with an *HBB* cDNA reference. The cDNA amplicon is aligning accurately to the reference and was the expected size of 251 bp.

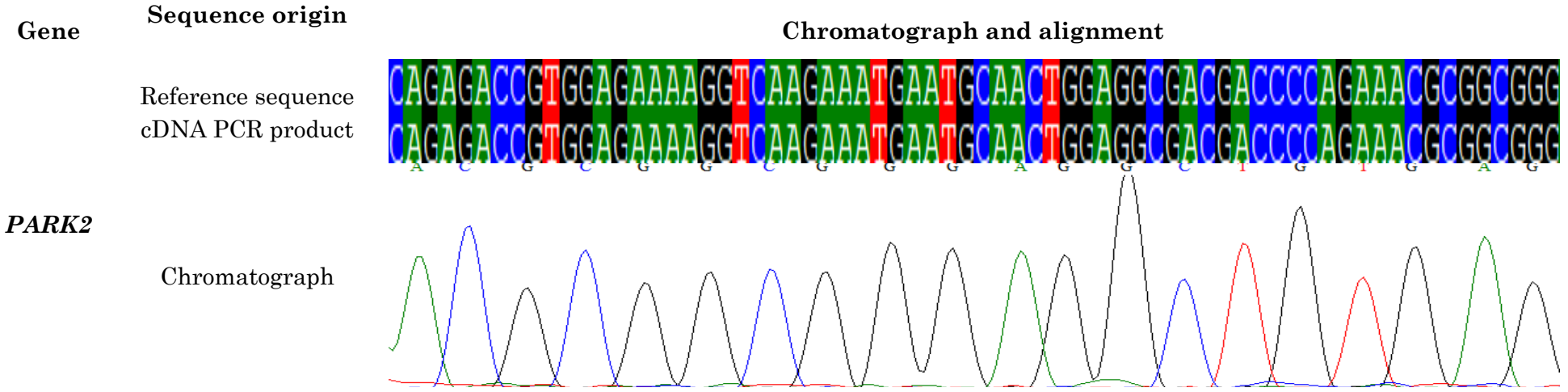


Figure 3.10 Alignment of *PARK2* cDNA fragment. The *PARK2* cDNA amplicon correctly aligned to the cDNA reference sequence which confirms that the correct product was amplified from the cDNA, it was also the expected size of 400bp.

3.11. Polymerase Chain Reaction of *NRXN2a* from cDNA

3.11.1. Amplification of Full-Length *NRXN2a* Inserts for Cloning

The reason behind generating this full-length *NRXN2a* insert was to clone the insert into a pDsRed-Monomer-C1 vector for transfection into SH-SY5Y cells for further study of the gene in a neuronal cell model.

Primers pairs (Section 2.10 in **Methods**), for the full-length *NRXN2a* gene were used to amplify full-length *NRXN2a* from the generated cDNA. The expected full-length *NRXN2a* fragment size is 5139 bp. In order to optimise the PCR conditions, annealing was performed at a range of temperatures from 50 to 62°C, at increments of 1°C (**Figure 3.11**). The gel electrophoresis showed that the majority of the 12 PCR reactions produced multiple bands and none of these bands corresponded to the expected size.

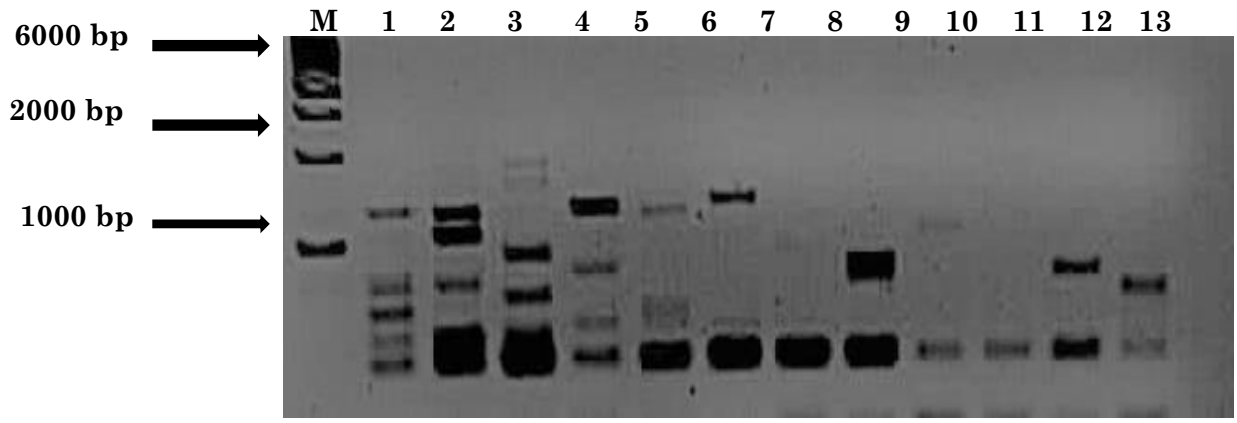


Figure 3.11 Visualisation of PCR products generated using full-length *NRXN2a* primers on a 1% agarose gel. In lane 1 is the lowest annealing temperature of 50 ^{0}C, and in lane 12 is the highest temperature of 62 ^{0}C. In the lanes between lanes 1 and 12, the annealing temperature increases in increments of 1^{0}C. The NTC is in lane 13. M is a 500bp DNA ladder (O'RangeRuler™ DNA Ladder, Thermo Scientific) and the expected PCR product size is 5139 bp.

The multiple products observed are likely as a result of non-specific primer binding. Further PCR optimisation was performed by adding dimethyl sulfoxide (DMSO) at a 5% concentration and the $MgCl_2$ concentration was also increased to 2,5 µM (**Figure 3.12)**. DMSO was added to the PCR because of the high GC content in *NRXN2a*, the rationale being that DMSO would forms hydrogen bond with template cDNA, and as a result destabilise the double helix structure. This should in theory allow for easier primer binding and full-length amplification. The increase of $MgCl_2$ concentration by a 0,5 µM increment was done to ensure verify that Taq polymerase had sufficient cofactor to

catalyse its activity and increase specificity. The resulting PCR products were larger than previously observed, between 2500 and 3000bp at annealing temperatures of 52 ^{0}C and 53^{0}C. However, a product of the correct size was not observed.

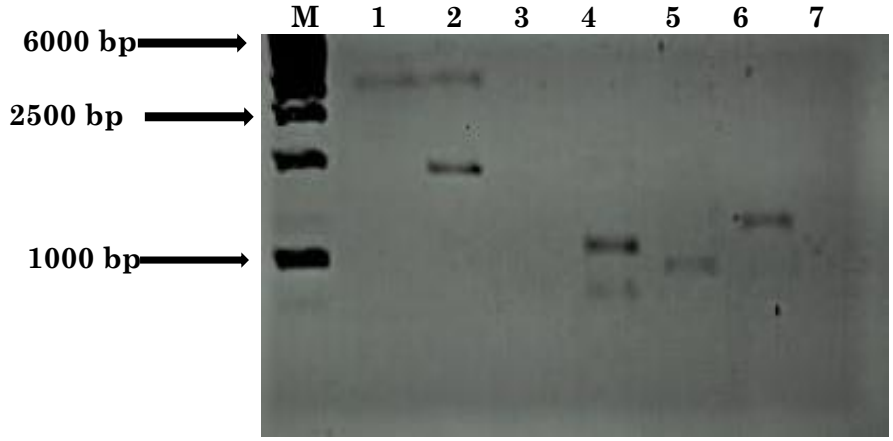


Figure 3.12 *NRXN2a* optimised PCR products on a 1% agarose gel. In lane 1 is the lowest annealing temperature of 52 ^{0}C, in lane 2 is 53 ^{0}C, in lane 4 is 54 ^{0}C, in lane 5 is 55^{0}C and in lane 6 is the highest temperature of 56^{0}C. M is a 500bp DNA ladder (O'RangeRuler™ DNA Ladder, Thermo Scientific) and the expected PCR product size is 5139 bp. The NTC is in lane 7.

In an attempt to generate the correct PCR product, a different polymerase was chosen for amplification. Instead of GoTaq DNA Polymerase, Phusion® High-Fidelity DNA Polymerase was used. Phusion DNA Polymerase is ideal for amplifying difficult and large amplicons because it has a DNA-binding protein domain which enables attachment to the template for very long amplifications. It also has proofreading ability, therefore it is likely to produce a product that is more accurate.

The gradient PCR was performed from 52 ^{0}C to 60^{0}C with increments of 1^{0}C (**Figure 3.13**). A single sized product was observed at all annealing temperatures with Phusion, showing greater specificity than that of the GoTaq reactions. The products amplified were much smaller (~2 kb) than the expected size, indicating that the desired product of ~5 kb was still not amplified.

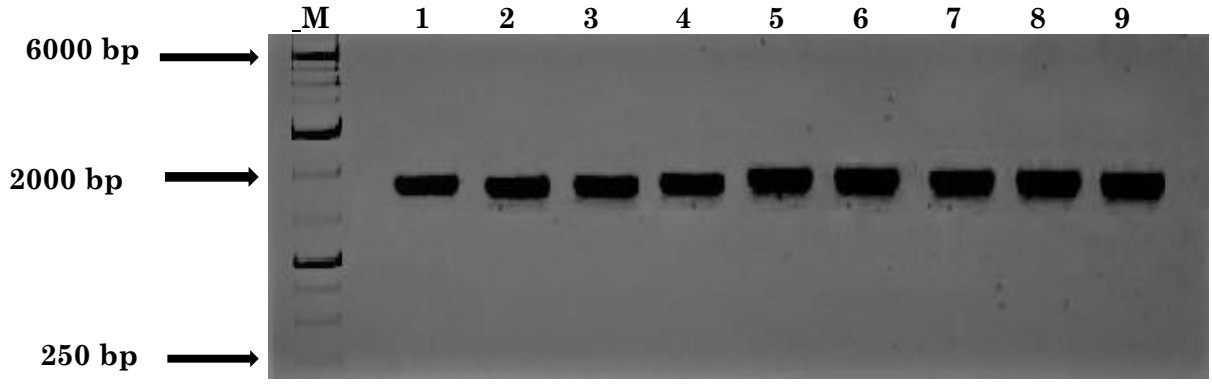


Figure 3.13 The amplification of *NRXN2a* using Phusion® High-Fidelity DNA Polymerase Taq, on a 1% agarose gel. The range of annealing temperatures is between 52 °C to 60°C, increasing by 1°C per lane. The PCR reaction are shown in lanes 1 to 9. A NTC was added to lane 10, which is not seen above. M is the1kb ladder .

3.11.2. Assessment of Polymerase Chain Reaction Using Large Products

To prove that the failure to amplify full-length *NRXN2a* from the cDNA was not due to experimental error or a problem with the PCR reagents, PCR amplification was performed using GoTaq DNA Polymerase on another similar-sized product.

The p2Nsm3125LS plasmid and PCR primers (Section 2.3, **Methods**) which are known to amplify a product of 4039 bp, were used. The reaction was done in triplicate and the expected PCR product was successfully amplified as shown in **Figure 3.14**. This showed that the failure to amplify the full-length *NRXN2a* gene was not as a result of technical or experimental problems, nor was it the size limitations of the polymerase or the reagents.

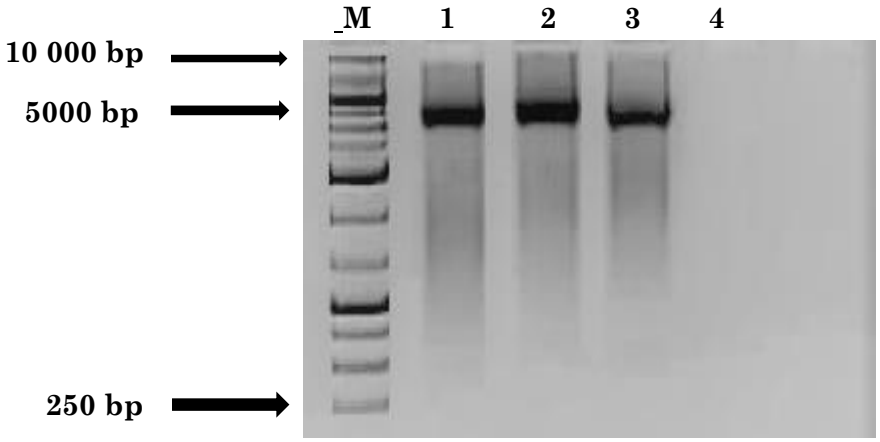


Figure 3.14 The amplification of the p2Nsm3125LS plasmid product shown on a 1% agarose gel. In lane 1 to lane 3 is the plasmid amplicon, in size it is just beneath the 5000 bp marker. M denotes the lane of the 1kb ladder used. The NTC is in lane 4

3.11.3. Polymerase Chain Reaction of *NRXN2a* Using Internal Primers to Generate Full-length CDNA

As an alternative to amplifying full-length *NRXN2a* cDNA (expected fragment size 5139bp) from the cDNA library generated, two new primer pairs were designed to amplify the *NRXN2a* gene in two overlapping fragments (one of 2593bp and the second of 2633 bp). These primers are described in **Table 2.2** in the Materials and Methods section. An internal NdeI site would facilitate subsequent 'stitching' of the two fragments to generate the full-length gene. Restriction enzyme sites for downstream ligation and cloning were added at the 5' and 3' ends of the gene. The rationale for these primers was that we thought that it may be easier to PCR the two smaller fragments instead of the 5kb fragment if the RNA in the library was partially degraded.

The initial attempts to amplify the two fragments using a standard protocol and PCR conditions were unsuccessful as seen by the lack of products in the lanes of the agarose gel electrophoresis in **Figure 3.15 A and B.**

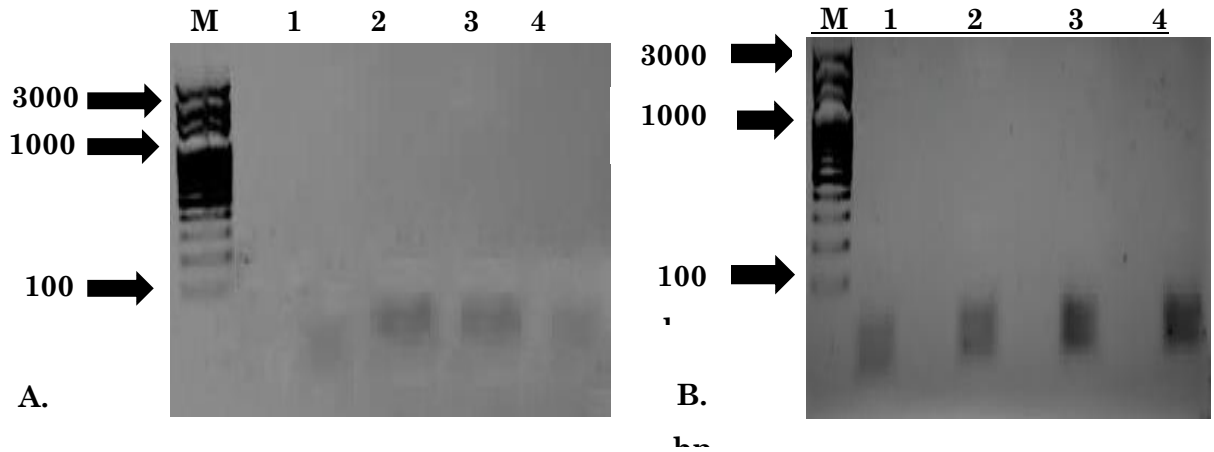


Figure 3.15 PCR amplification of shorter *NRXN2a* primers on a 1 % agarose gel. **A.** *NRXN2a* EcoR1 5' Forward and *NRXN2a* Internal 2 Reverse, the expected PCR product size is 2594 bp. In lane 1 to lane 3 are the three cDNA amplification products done in triplicate. The NTC is in lane 4. **B.** *NRXN2a* Internal 1 Forward and *NRXN2a* SAL1 3' Reverse, the expected PCR product size is 2633bp. In lane 1 to lane 3 are the three cDNA amplification products done in triplicate, the NTC is in lane 4 and M is a 100 bp DNA ladder (Solis BioDyne).

Subsequently, we attempted to optimise the PCR by changing various factors. These included changing from Taq DNA Polymerase to High-Fidelity DNA Polymerase, the use of different annealing temperatures, modifying the concentration of the $MgCl_2$ from 1,5 µM to 2,5 µM and increasing the extension according to product size. However, the

optimisation attempts on the shorter *NRXN2a* primers to amplify the *NRXN2a* gene in two separate fragments were unsuccessful.

3.11.4. cDNA Synthesis Using a *NRXN2a* Specific Reverse Transcription Gene Primers

Lastly, in an effort to increase the reverse transcription efficiency and target specificity of the *NRXN2a* gene during cDNA synthesis from the Human Foetal Brain RNA library, we decided to use *NRXN2*α-specific primers during the initial cDNA synthesis step. A combination of four primers were designed that would bind to the 3´untranslated region (UTR) of the *NRXN2a* RNA, and these would be used instead of the random hexamers that were supplied with the QuantiNova™ Reverse Transcription kit, for cDNA synthesis.

Due to cDNA being synthesised from multiple target specific primers, it is expected that if the resulting product was loaded on a gel there would be cDNA fragments around the size of the target gene (5139 bp). The gel electrophoresis (**Figure 3.16**) of the 'gene specific' cDNA shows a faint smear. The cDNA in lane 1 was previously synthesised from the RNA library using the hexamer primers that came with the kit and it was loaded as a positive control for reverse transcription.

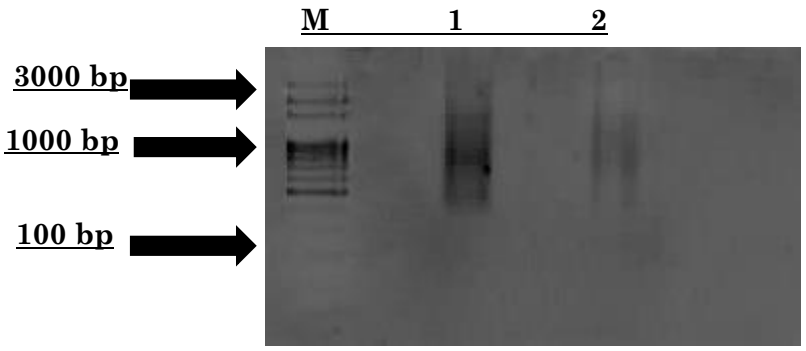


Figure 3.16 *NRXN2a* gene specific primers cDNA synthesis. Agarose gel electrophoresis of cDNA generated from a Human Foetal Brain RNA library. In lane1 is the previously generated cDNA using the hexamer primers included in the reverse transcription kit. In lane 2 is cDNA synthesised using *NRXN2a* specific primers. M is a 100 bp DNA ladder (Solis BioDyne).

Next, we attempted to amplify from this gene specific cDNA the *NRXN2a* gene using the primer sets that produce two overlapping fragments of ~2,5 kb. Standard PCR protocol and conditions were used in a gradient PCR to use multiple annealing temperatures on the primer sets. Though the gene specific cDNA was generated, the PCR of the two overlapping fragments of ~2,5 kb using normal Taq polymerase was unsuccessful. The PCR was repeated using the QIAGEN LongRange PCR Kit on the gene specific cDNA, but no amplification occurred.

In summary, despite the design of several primer pairs, changing PCR reagents and multiple rounds of PCR optimisation attempts with both the hexamer generated cDNA as well as the 'gene specific' cDNA, full-length or two overlapping fragments of *NRXN2α* were never successfully produced from the Foetal brain library.

3.11.5. Intron Spanning Primers to Generate Smaller Fragments of NRXN2α cDNA

Since amplification of both full-length *NRXN2α* (~5kb) and the two overlapping fragments of *NRXN2α* from the gene specific cDNA were unsuccessful, we then considered whether the problem was the *NRXN2α* expression in the Human Foetal Brain RNA library. This could be as a result of degradation of the mRNA library caused by repeated freeze thaw cycles over time, or low levels of expression in the tissue from which the library was prepared.

To examine this, we designed PCR primers to span small regions of *NRXN2α* cDNA. If a PCR product is produced it would mean that the synthesised cDNA library contained *NRXN2α* gene DNA fragments spanning multiple exons. The first primer set designed spanned from exon 2 to exon 7 and would produce a product of 475bp. The second primer set spanned from exon 19 to exon 22 and would produce a PCR product of 515bp.

After optimisation, both the exon spanning primers successfully amplified the expected PCR product from the cDNA**.** The product generated from the exon 2 to exon 7 primer set produced a band close to 500 bp in size on the gel (**Figure 3.17 A**). The product was subsequently Sanger sequenced and produced the correct sequence which was 450 bp (**Figure 3.18)** with a sequence identity score of 92% to the *NRXN2*α cDNA reference. The exon 19 to exon 22 primer set, produced a band that was slightly above 500 bp in size (**Figure 3.17 B**). This product was also Sanger sequenced and produced a sequencing product that was 527 bp (**Figure 3.19)** and had a sequence similarity score of 88% compared to *NRXN2α* cDNA reference. The observed variation between the sequenced products and the gene reference might explain why the amplification of full-length *NRXN2*α did not work, primer binding sequences found using the reference might be not be locatable in the cDNA.

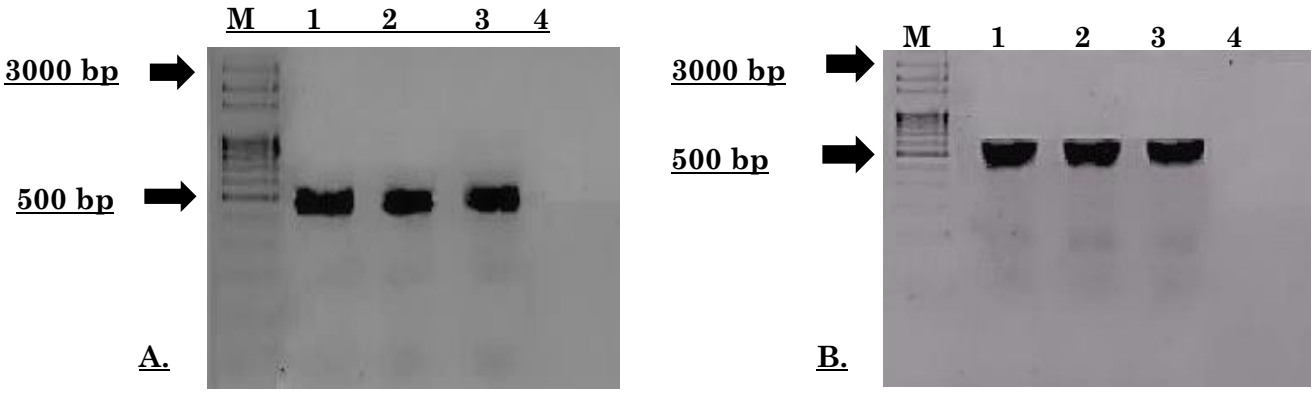


Figure 3.17 Gel electrophoresis of PCR products using the primers to generate smaller fragments of *NRXN2a*. The PCR products are visualised on a 1% agarose gel. A. *NRXN2a* Exon 2 to 7 product and the expected size is 475 bp. B. The gel of *NRXN2a* Exon 19 to 22 product and the expected size is 505 bp. In both gel image A and gel image B lane 1 to lane 3 are the gene specific cDNA amplifications products done in triplicate, the NTC is in lane 4 and M is a 100 bp DNA ladder (Solis BioDyne).

The successful amplification of the fragments proves that there are *NRXN2a* fragments in the generated cDNA from the Human Foetal Brain RNA library. Sanger sequencing validated the products. This could mean that the gene cDNA might be fragmented, allowing only the amplification of smaller fragments. This might explain why the full-length primers are unable to bind to the ends of the gene.

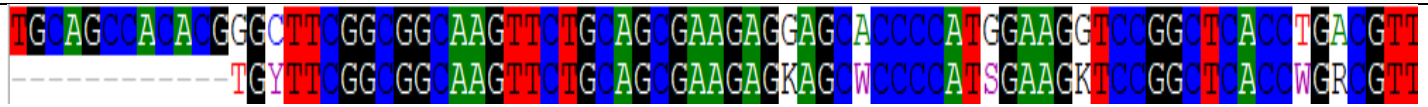
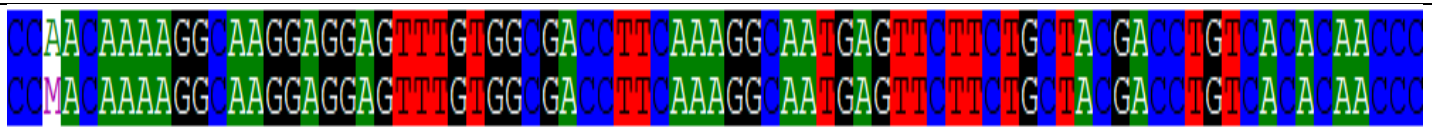
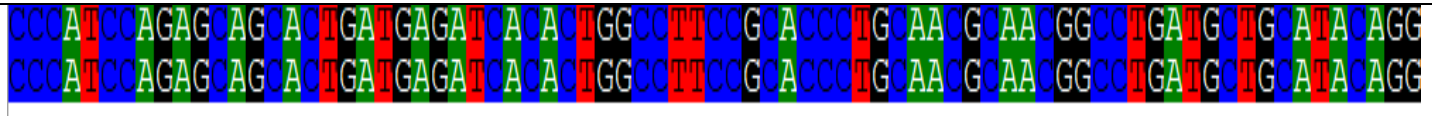
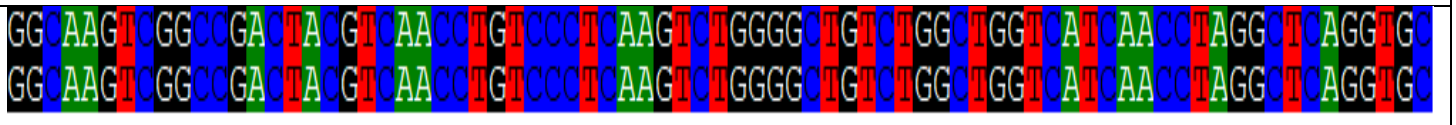
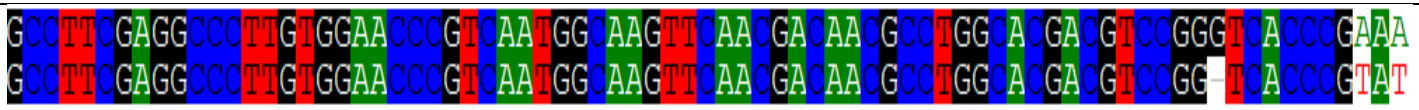
Sequence origin	Bp	Alignment
Reference sequence cDNA PCR product	1 – 75	
Reference sequence cDNA PCR product	75 – 150	
Reference sequence cDNA PCR product	150 – 225	
Reference sequence cDNA PCR product	225 – 300	
Reference sequence cDNA PCR product	300 – 375	
Reference sequence cDNA PCR product	375 – 450	
Reference sequence cDNA PCR product	450 – 525	

Figure 3.18 Sequence chromatograph and alignment Exon 2 to exon 7 cDNA product. The reference used for the alignment contains a 710 bp exon spanning from the end of exon 1 to the beginning of exon 8 region. The sequencing confirmed that the amplified product was the correct sequence and size.

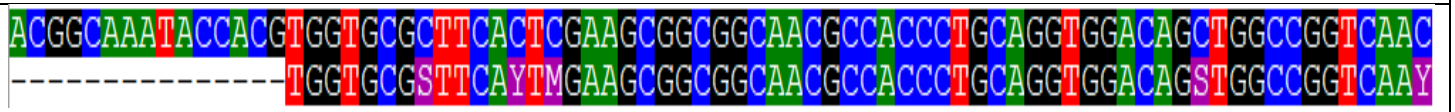
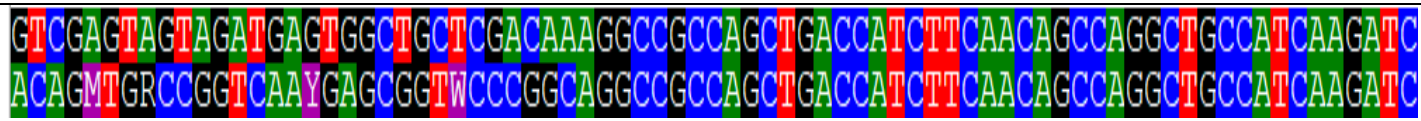
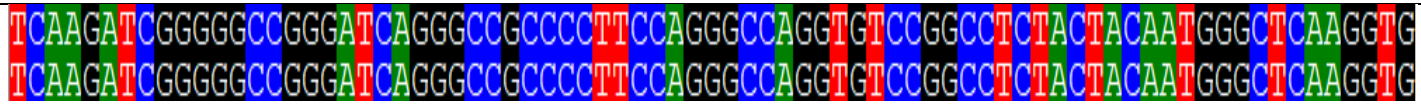
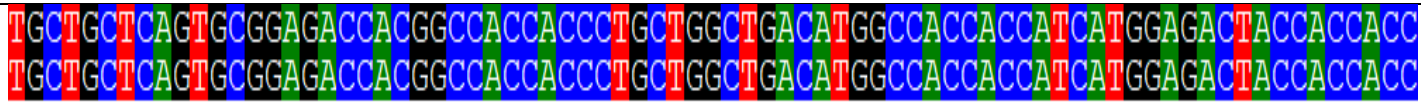
Sequence origin	Bp	Alignment
Reference sequence cDNA PCR product	1 – 75	
Reference sequence cDNA PCR product	75 – 150	
Reference sequence cDNA PCR product	150 – 225	
Reference sequence cDNA PCR product	225 – 300	
Reference sequence cDNA PCR product	300 – 375	
Reference sequence cDNA PCR product	375 – 450	
Reference sequence cDNA PCR product	450 – 525	
Reference sequence cDNA PCR product	525 – 600	

Figure 3.19 Sequence chromatograph and alignment of Exon 19 to exon 22 cDNA product. The reference used for the alignment contains a 660 bp exon spanning from the end of exon 18 to the beginning of exon 23 region. The sequencing confirmed that the amplified product was the correct sequence and size.

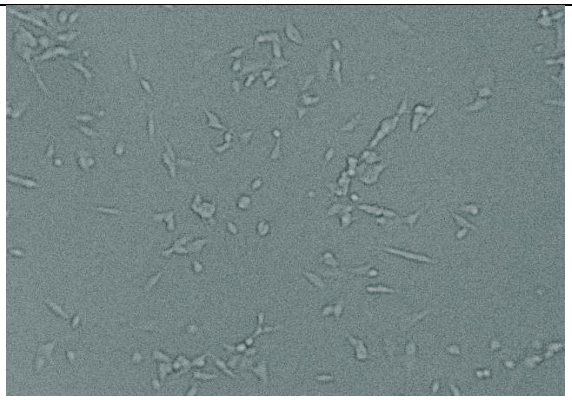
3.12. NRXN2α Protein Analysis

3.12.1. Determining the NRXN2α Protein Concentration in Neuroblastoma Cells

It was initially the intention to study the functional effect of NRXN2α G849D in comparison to the wild-type in an appropriate neuronal cellular model. Due to the inability to generate the full-length cloning insert of *NRXN2α* for the functional effect studies, we instead evaluated the Neuroblastoma (SH-SY5Y) cell line as a suitable *in vitro* model for future NRXN2α expression studies. To evaluate the suitability of this model, we determined the expression of endogenous NRXN2α protein in these cells and tested the reproducibility of the NRXN2α antibody in Western blotting experiments.

3.12.1.1. Cell Proliferation

Undifferentiated SH-SY5Y cells were cultured in medium containing all the essential components for the mitogenic effect (proliferation and cell growth). The cells grow as clusters of cells with multiple, short, fine cell processes (**Figure 3.20**). We simultaneously cultured four passage-12 SH-SY5Y cell flasks with a surface area of 75 cm^2, for 5 days before lysing on the sixth day of proliferation. The majority of the cells were adherent and there was no significant morphological differentiation of SH-SY5Y cells over the culturing time. The images below are representative pictures of the growth rate taken from only one of the four flasks.

Days in culture	Cell image
Day 1 (post thaw) Low Density	

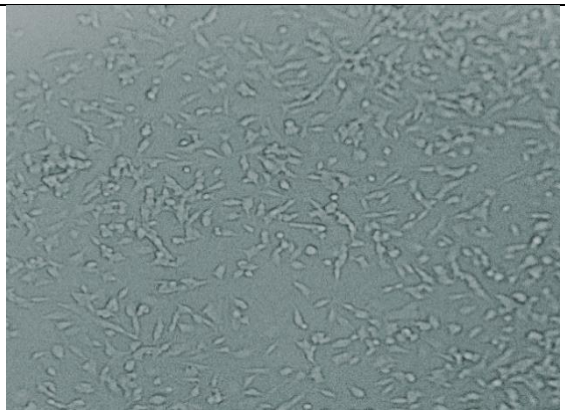
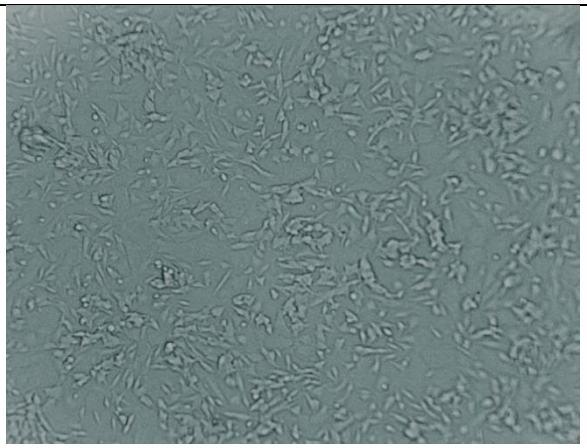
Day 3 70 % confluent	
Day 5 90% confluent High Density	

Figure 3.20 SH-SY5Y cell line proliferation. The cells which were incubated in growth media show an adherent growth pattern. The cells were examined in a Leica DMI 6000B inverted microscope and photographed at 20 and 100 x magnification. Data are representative pictures from one of four sample flasks.

3.12.2. Western Blot Analysis of Endogenous NRXN2α Expression

Firstly, aliquots of the protein lysates were used to determine the protein concentration by comparison to known standard concentrations of BSA ranging from 0-1000ng/ml. This was done using the colorimetric Bradford assay (**Figure 3.21**). The assay determines protein concentrations using the sample absorbance read at 595nm. The BSA standard curve was generated and the value of the slope of this standard curve, calculated from the line of best fit. This was used to calculate the concentration of the SH-SY5Y protein lysates. Four protein lysate samples were made from each of the four flasks and kept separate throughout the experiment. The protein concentrations were measured in triplicate and this provided an estimate of protein concentration in µg/µl which was used to calculate the volume required to load 100µg of sample into individual lanes of the SDS gel for the Western blotting experiments.

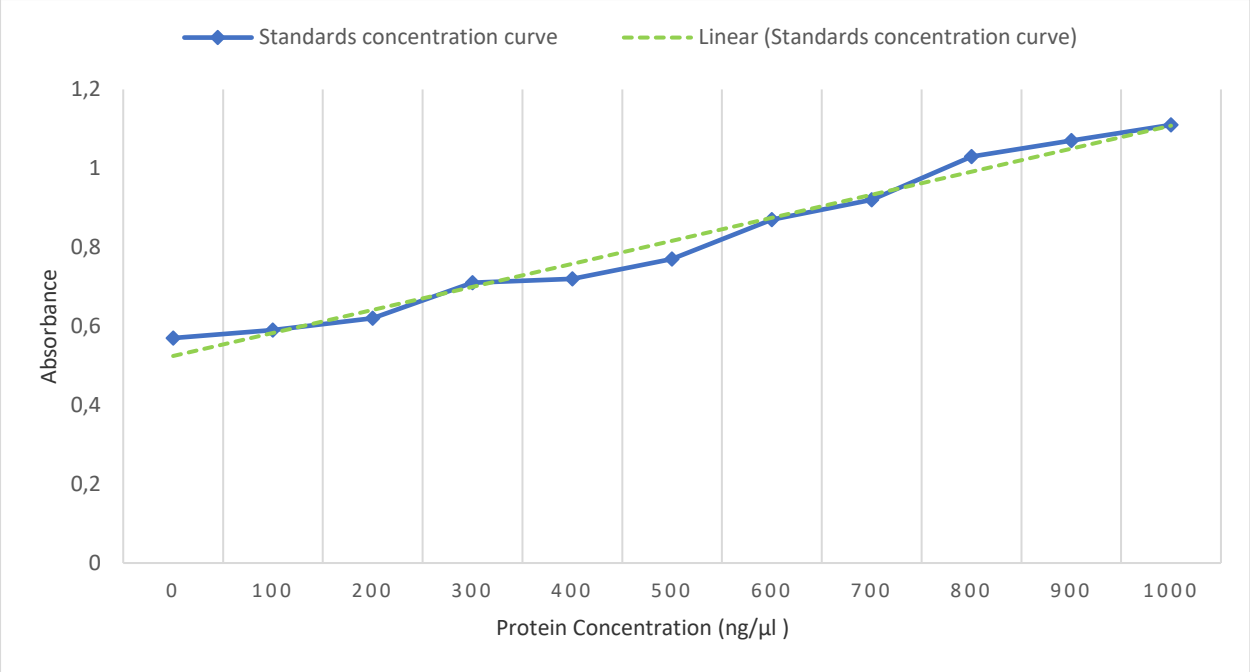


Figure 3.21 Protein assay standard curve. A standard curve (in blue) generated from BSA standards of 0 ng/µl, 100 ng/µl, 200 ng/µl, 300 ng/µl, 400 ng/µl, 500 ng/µl, 600 ng/µl, 700 ng/µl, 800 ng/µl, 900 ng/µl and 1000µg. The absorption of each standard was measured at 595nm. A line of best (in green) was added and the value for the slope of this line was used to calculated from the concentration of protein samples.

Western blots were performed on the four protein lysates. The protein concentrations of the lysates were measured in triplicate and this provided an estimate of protein concentration in µg/µl which was used to calculate the volume required to load 100µg of sample into individual lanes of the SDS gel for the Western blotting experiments. The optimal dilution of each of the antibodies was then experimentally determined by varying the dilution factor based on the recommended dilution. The Anti- NRXN2α antibody was diluted 1/400 (recommended), 1/500 and 1/1000 (**Table 2.7**, Methods). The Anti-GAPDH antibody was diluted 1/1000 (recommended) and 1/3000. The optimal dilutions for the Anti- NRXN2α antibody was found to be 1:500 (**Figure 3.22**), at 1/1000 no protein was visible. For the Anti-GAPDH antibody the optimised dilution was 1/3000. Using the optimised antibody dilutions, the experiment was repeated in triplicate (n=3). Anti-NRXN2α antibody at all dilutions showed bands at about 270 kDa, this did not correspond to the expected size of the endogenous protein (185 kDa). No non-specific binding or background was visible on the blots. A 37 kDa band for GAPDH was seen as expected.

The size difference between the observed band 270 kDa and the expected band size 180 kDa could be a result of incomplete unfolding which would influence electrophoretic mobility, preventing the protein from moving in the gel. Another possibility is that NRXN2α is a transmembrane protein, which are known to tend to oligomerize and

aggregate when heated. Heating samples to 95°C is a vital in the standard denaturing process before loading onto the SDS gel. Another reason we might be seeing the larger NRXN2α molecule could be because it is bound to another protein such as Neuroligins which are typically about 100 kDa in size. Protein identification using LC-MS-MS could be used to confirm that this band corresponds to NRXN2α and if it is bound to another protein it can also assist in its identification.

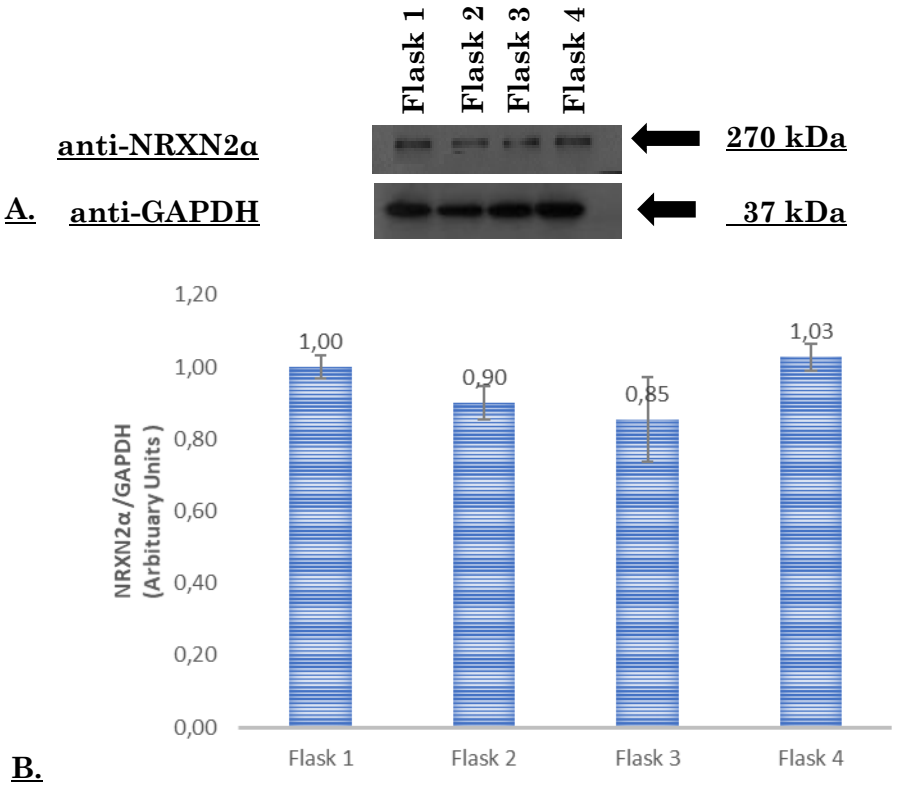


Figure 3.22 Western blot showing endogenous NRXN2α expression in proliferating in SH-SY5Y cells. A is a representative blot showing the expression of endogenous NRXN2α protein in SH-SY5Y cells protein lysates. The 270 kDa band represents the NRXN2α protein detected by a 1:500 anti-NRXN2α antibody. Cytosolic marker, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is used as a loading control, seen at approximately 37 kDa. Single representative blot is shown (n=3). B is a quantitative assessment showing the relative amount of endogenous NRXN2α protein in the protein lysates. NRXN2α was expression was measured as a ratio to GAPDH expression. Error bars indicate standard deviation of three experiments. Quantitative measurements of bands were performed using densitometry (Fiji software).

A quantitative assessment of the blots was performed to measure and normalise the relative amount of endogenous NRXN2α expression to the loading control. For the blots with the 1:500 anti-NRXN2α antibody dilution, NRXN2α was stably expressed across all four flasks (**Figure 3.22**). The aim of these Western blotting experiments was to evaluate

the SH-SY5Y cell line as a suitable *in vitro* model for the functional study of NRXN2α expression. The experiments suggest that undifferentiated, SH-SY5Y cells that had not been transfected do express endogenous NRXN2α, however this needs to be verified using other methods.

Chapter Four

4. Discussion

There have been numerous studies that have established a genetic aetiology for PD in non-African populations but to date, studies on SSA patients have shown that the established pathogenic mutations in PD genes are a minor contributor (Okubadejo et al., 2008; Keyser et al., 2009; Bardien et al., 2009; Haylett et al., 2012; Yonova-Doing et al., 2012). The present study, which was an effort to investigate a multiplex family with PD in order to identify a pathogenic mutation in a new PD gene or a novel mutation in a known PD gene, aims to contribute to this knowledge gap. For the realisation of the study objectives, the methodology employed included WES, Sanger sequencing, bioinformatic tools, HRM analysis, cell culture and Western blotting.

4.1. WES Results and the Prioritisation Process

Here, we report the WES analysis conducted on a SA Afrikaner family exhibiting an AD form of PD with multiple affected individuals. Based on the absence of any of the known PD mutations in this family, as determined by WES conducted on three of the affected individuals (11.844,12.726 and 92.32), it was concluded that a novel mutation was responsible for the disease.

As anticipated, the number of variants per individual yielded by WES exceeded 22 000, requiring the application of a stringent filtration criteria. WES analysis is a useful tool in the diagnostic screening of known disease-causing genes (Yang et al., 2013), however, it requires either large families (with multiple affected and unaffected individuals), for which genotype–phenotype co-segregation data can be obtained, or ,large case-control cohorts, in order to confirm any positive findings (Vilariño-Güell et al., 2011).

The WES quality control was conducted by sequence alignment, variant calling, and annotation. Quality, quality measures (coverage, read depth) and population frequency filters were used to remove any variants that were commonly found in the population or variants of low sequencing quality, as they would likely be artefacts. Next, all the synonymous co-segregating variants were excluded from the list of candidates as the relevant variant was expected to modify the encoded protein.

Considering the AD pattern of inheritance, the next criteria applied was that the variants must be heterozygous and must be shared between all the affected individuals, this reduced the number of variants significantly from nine to five.

Four of the five validated variants are novel (*RFT1* p.A463G*, NRXN2* p.G849D*, TEP1* p.Y412C and *CCNF* p.C363S*)* and one variant in *CFAP65* p.T1023A had been previously reported (https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgirs=rs748008106). In an effort to further prioritise the candidate list, the five variants were assessed for co-segregation amongst the four unaffected family members, and the two newly diagnosed family members (92.33 and 11.902). Considering the dominant inheritance pattern of disease, any variant that was present in the unaffected family members could be excluded as this would be strong evidence that the given variant is not disease-causing (Farlow et al., 2016a; Ng et al., 2010; Verstraeten et al., 2015). The *CCNF* p.C363S variant was found in three unaffected individuals, this excludes it as the potential causal candidate and indicates that it might be a rare polymorphism.

Four candidates co-segregating only in the affected individuals were therefore identified namely, *RFT1* p.A463G, *NRXN2a* p.G849D, *TEP1* p.Y412C and *CFAP65* p.T1023A. Of the two newly diagnosed PD patients only one individual (92.33; mother of affected individual 92,32) shared all five validated variants with the three individuals on whom WES was performed. The other recently diagnosed PD patient (11.902) had only one variant in common with the other four affected individuals, which was the previously reported *CFAP65* p.T1023A variant. It should be noted that four of the affected individuals all exhibited typical to mild PD, while individual 11.902 presented with a very late AAO with prominent symptoms being tremor and REM sleep behaviour disorder. This individual might be a phenocopy of disease as she is both genetically and phenotypically distinct from the other four.

The disease presentation of a phenocopy is usually similar to the disease of other individuals of the same family. However, the genotyping of that phenocopy would elucidate the absence of the mutation(s) responsible that are found in the 'true affected' individuals who are mutation carriers (Nichols et al., 2005; Klein et al., 2011). Phenocopies in PD are well-documented, in a study of multiplex PD families by Nichols et al., (2005) with *LRRK2* p.G2019S affected mutation carriers , it was found that approximately 25% of the affected siblings to the mutation carriers were homozygous for the wild-type allele, making them phenocopies of PD. In a familial PD study by Klein et al., (2011) conducted on 23 multiplex PD families, 27 PD patients (14,4%) were found to be phenocopies (Klein et al., 2011). As a result, individual 11.902 was considered to be a phenocopy, which is defined as an environmentally induced disease phenotype or that is genotypically different from the actual cause of disease.

To distinguish between what could be population specific polymorphisms and disease-causing variants, we screened the four variants by HRM in 192 ethnically-matched neurologically normal controls and found that all four variants were absent in the 384 screened alleles. Further screening of the four variants in publicly available databases ExAC and the 1000 Genomes Project, showed that the three novel variants (in *RFT1, NRXN2a* and *TEP1)* were absent in multiple populations (**Table3.6**). The previously reported *CFAP65* variant, was however found in the ExAC database at a MAF frequency of 17 out of 2 100 000 individuals (give allele frequency) in the ExAC Aggregated Population (https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=rs748008106. Although this variant was found in an online database which may include some diseased individuals, its MAF was rather low and additional information was required before it could be ruled out as the potential causal variant.

To add further information about the potential causal role of variants before prioritising, various tools were applied. First, functional prediction tools (SIFT, PolyPhen-2, MutationTaster and CADD) were used to predict the likelihood of functional impact on the protein, and GERP which denotes the level of evolutionary constraint acting at a specific site in the genome was used. Subsequently, tissue expression data and pathway analysis were determined using publicly available databases to ascertain where the gene was highly expressed and in which pathways. None of the four variants were found to be either co-expressed, co-regulated or interact with any of the known PD genes.

The factors, which initiated and later strengthened the suspicion of the causative nature of the *NRXN2a* p.G849D variant were the function and expression data of the protein, the predicted deleterious effect of the amino acid changes to the protein's function, the segregation of the missense variant in only affected individuals, the conservation of the area where the variant is located and the low frequency of the novel variant in both online databased and ethnically-matched controls. Based on meeting the filtration criteria and being the most relevant to PD pathology, *NRXN2a* was selected as the top candidate for further study.

The co-segregating variants shared between the three affected individuals were narrowed down to a list of candidates in a step-wise manner, using a comprehensive filtration approach. In recent years this method of filtration has been successfully applied in multiple studies investigating PD gene discovery. Most of the studies considered the following criteria for filtration and prioritisation (1) where the variant occurs (exonic, intronic or splice region), (2) the allele frequency of less than 1% in the control populations

of annotated frequency databases, (3) were predicted to be damaging by multiple in-silico functional and structural prediction programs, (4) be located in a highly conserved region, (5) were in a relevant Gene Ontology (GO) category, (6) were expressed in the brain, (7) segregated in at least 2 PD cases of the same family, and (8) were absent in the unaffected members of the families. The PD gene candidates that have been identified using this filtration approach include *Parkin* (p.Q34R and (p.V380L) (Myhre et al., 2008), *CHCHD2* (p.T61I) (Funayama et al., 2015), *TMEM230* (p.184ProGext*5, p.R141L) (Deng et al., 2016; Yan et al., 2017; Yang et al., 2017), *TNK2* (pA977V) variant and the *TNR* (pT166A) (Farlow et al., 2016).

In 2015 the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published standards and guidelines in order to establish consistency in variant classification among clinical genetic testing laboratories (Richards et al., 2015). These guidelines describe a framework for classifying variants as "pathogenic", "likely pathogenic", "uncertain significance", "likely benign" or "benign" according to a series of criteria with levels of evidence defined as very strong, strong, moderate or supporting (Richards et al., 2015). The criteria for classification according to these standards is

1. The information available on population frequency databases like ExAC and dbSNP
2. The predictions made by multiple computational prediction tools
3. Previously published functional studies showing the impact of modifying the protein
4. The variant segregation data attained from genotyping the family
5. Allelic Data, how the variant is inherited (heterozygous, homozygous, dominant or recessive)
6. The clinical information that is available on the patients (Optional category as it's not always available in the laboratory setting)

According to the evidence classification of the ACMG guidelines the *NRXN2a*. p.G849D mutation is considered pathogenic with strong evidence. It gained a pathogenic moderate 2 rating for frequency in population databases. The rating for computational prediction tools evidence was pathogenic strong 1. The rating gained from the information on the gene in previously published functional studies was pathogenic moderate 1. The segregation evidence got a pathogenic strong 1 with supporting evidence. Allelic data

criteria (heterozygous, with dominant inheritance) got a pathogenic moderate 2 rating. Lastly because we had clinical data of the family available that informed the diagnosis and described the disease in each individual, we gained a pathogenic with supporting evidence 4 rating.

The filtration strategy that has been used in this study is in line with the ACMG guidelines going beyond what is suggested in the guidelines by obtaining additional information such as the protein expression data in the relevant tissue, structurally modelling the affected protein, screening the local control population and additionally screening other patients affected by the disease.

4.2. Neurexin Protein Family

Neurexins (*NRXNs*) constitute a family of highly polymorphic transmembrane proteins, reported to be highly expressed in the brain (Schreiner et al., 2014, 2015). They were initially identified at the synapse as calcium-dependent a-latrotoxin receptors (Ushkaryov et al., 1992; Ushkaryov and Südhof, 1993). *NRXN1* [OMIM #600565] and *NRXN3* [OMIM #600567] from this family of genes are amongst the largest genes in the mammalian genome, exceeding 1,6 Mb, and each gene encodes a longer α and a shorter ß transcript. Due to multiple promoters and alternative splicing, thousands of isoforms are generated from the three neurexin genes (*NRXN1*, *2* and *3*) (Missler and Südhof, 1998). Though these NRXN isoforms are co-expressed, each has a different mRNA and resulting protein. Furthermore, they are differentially distributed in neuronal cell types according to their specific function (Ullrich et al., 1995; Schreiner et al., 2014, 2015).

Structurally αNeurexins are composed of six large extracellular laminin/neurexin/sex-hormone binding globulin (LNS) domains, with three interspersed epidermal growth factor (EGF)-like regions as seen in **Figure 4.1**. Intracellularly, all neurexins contain a transmembrane region, a carboxy terminus and a short intracellular tail. The shorter ß neurexins only have the sixth LNS domain on the extracellular side, and no EGF-like regions (Missler and Südhof, 1998; Rudenko et al., 1999, 2001).

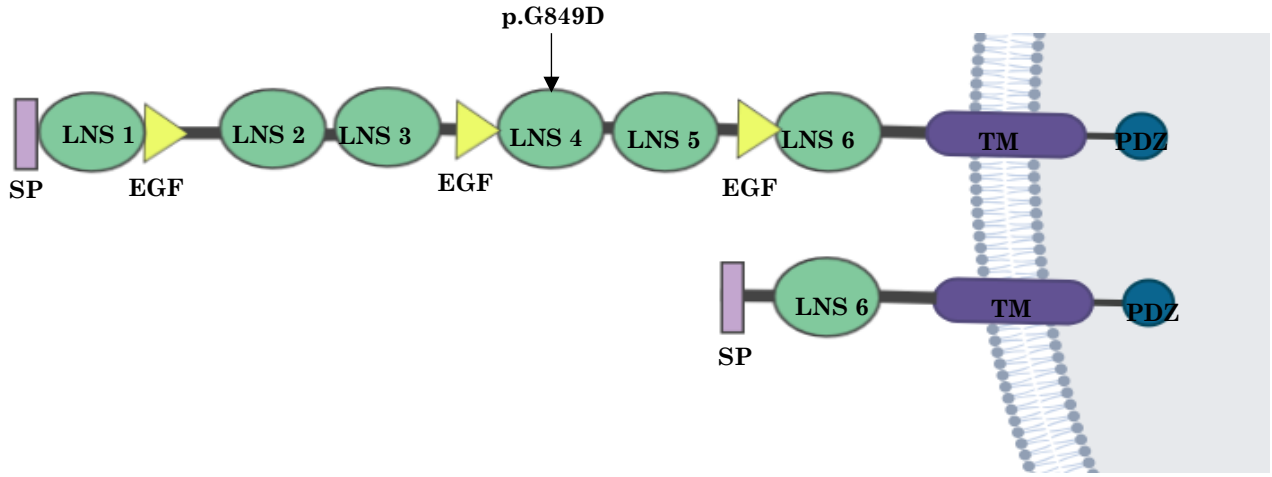


Figure 4.1 Schematic structure of the two major isoforms (alpha and beta) of the NRXN2 proteins. SP, signal peptide; LNS 1–6, laminin neurexin sex hormone binding domains; EGF, epidermal growth factor-like domains; TM, transmembrane region; PDZ, domain binding site. The Black arrow indicates the location of the missense variant p.G849D found in the affected patients of this study. Image generated using Biorender.

Studies have shown that neurexins are embedded in the presynaptic membrane and interact with multiple key postsynaptic surface proteins (**Figure 4.2**) namely Neuroligins (NLGNs) (Craig and Kang, 2007; Mukherjee et al., 2008), Leucine-rich repeat transmembrane neuronal proteins (LRRTMs) (de Wit et al., 2009; Ko et al., 2009; Siddiqui et al., 2010), Cbln (1 precursor protein) and GluR*δ* (δ-type glutamate receptor) (Uemura et al., 2010). Thus, neurexins are emerging as a very important family of synaptic regulation proteins. **Figure 4.2** shows the schematic representation of the long NRXN2α and the shorter NRXN2β binding to NLGN in the synaptic cleft through the sixth LNS domain on the extracellular side of the neuronal membrane. The prioritised gene, *NRXN2α* [OMIM #600566], is known to be involved in calcium channel regulation, transmembrane signalling receptor activity, neuronal cell adhesion, synaptic organization and neuroligin family protein binding (Craig and Kang, 2007).

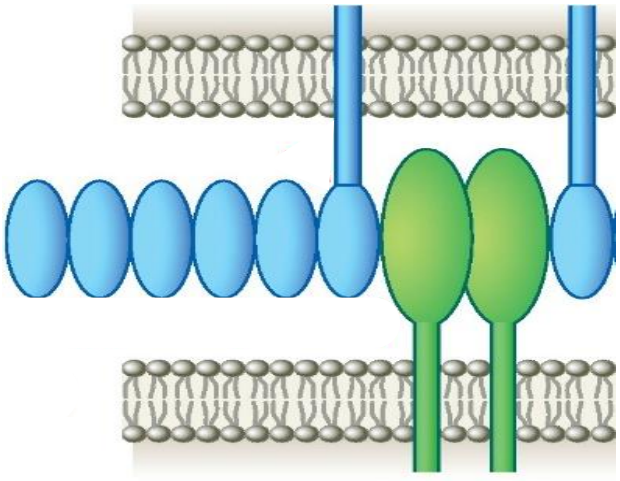


Figure 4.2 Schematic representation of NRXN2 and NLGN complex formation at the synaptic cleft. NLGN is shown in green. NRXN*α* and NRXN*β* are shown in blue. The NRXNs bind to NLGN using the sixth LNS domain extracellular domain. Used with permission from Elsevier.

4.3. Involvement of Neurexins in Human Disorders and Animal Studies

Of the three neurexins, only NRXN1 inactivation has been well studied and it has been associated with multiple neurodevelopmental disorders including schizophrenia, Tourette syndrome, epilepsy and Autism Spectrum Disorders (Schaaf et al., 2012; Møller et al., 2013; Yangngam et al., 2014; Dachtler et al., 2014; Chen et al., 2017; Dachtler et al., 2015).

It is only recently that NRXN2 inactivation has come into focus in a study investigating synaptic impairment in mice (Dachtler et al., 2014). The inactivation of both NRXN1 and NRXN2 caused the dysregulation of neurotransmitter release, however the synaptic trafficking was not completely lost (Dachtler et al., 2014). Although these results do not directly explain the direct contribution of neurexins to Autism pathology, the study concluded that the observed deficiencies in presynaptic vesicular release confirmed the pathological significance of the neurexins in neurological disorders (Dachtler et al., 2014).

In animal studies, genetic mutations of each neurexin uncovered diverse phenotypes that affect multiple aspects of synaptic function. In mice, constitutive single knockout (KO) of NRXN1α or NRXN2α with the expression of ß -neurexins severely impaired the survival of the mice. Subsequently, the KO of all of the α -neurexins was invariably lethal (Missler et al., 2003). The α -neurexins KO mice had a decrease in inhibitory but not in excitatory synapse activity and displayed a dramatic impairment synaptic transmission, this was hypothesised to be as a result of a loss in presynaptic Ca2+ influx (Missler et al., 2003). Similarly, triple ß -neurexin KO mice also exhibited a synaptic phenotype, however it was milder due to the compensatory mechanisms of the α -neurexins (Anderson et al., 2015;Fuccillo et al., 2015)).

In another animal study by Tabuchi and Südhof, (2002) a mutation was created in the mouse NRXN2*α* cDNA, at the end of the fourth LNS domain. This did not affect NRXN2*β*, but about half of the NRXN2*α* protein was lost. The region lost included the binding sites for NLGNs and LRRTMs which are located in LNS 6. Furthermore, the NRXN2α mutant exhibited a loss of synaptic activity in the neuron co-culture assay. In the presence of the wild-type NRXN2α clustering of glutamatergic components LRRTM2, PSD-95, NLGN2 and gephyrin at dendrite contact sites was observed. The mutant however failed to cluster these postsynaptic components (Tabuchi and Südhof, 2002).

In summary, the genetic alternations that have been performed on *NRXN2α* showed a complete loss of function in binding to Ca^{2+}, LRRM2 or NLGN2 in cell-based assays, as well as a complete loss of synaptic activity in animal models. It is therefore plausible that the *NRXN2α* gene may be involved in a neurological disease like PD.

4.4. Possible Effect of the Missense Variant in *NRXN2α*

All of the functional prediction tools had predicted that the missense variant in *NRXN2α* is likely to have a damaging effect on the protein's function. The variant results in the exchange of a smaller non-polar amino acid for a larger negatively charged amino acid with a carboxyl group side chain (CH2COOH), in a prominent loop of the LNS 4 domain.

This will potentially result in the disruption of the protein structure as the negatively charged amino acid may interact with the neighbouring residues which may stabilize what should be a flexible conformation of the LNS domain, meant for optimal extracellular binding. As previously discussed, the missense variant is located in a highly conserved area of the protein. This high degree of conservation potentially implies that this region of the protein may have important functions. Based on the prediction of high conservation, if this region is affected by the variant, it could lead to the instability of the protein.

The extracellular region of the NRXN2α protein mediates synaptic protein interactions, cell adhesion, cell recognition and synaptic regulation. As a result it can speculated that the disruption of NRXN2α might affect the protein binding properties, synaptic transmission efficacy, the NRXN2α receptor function, presynaptic vesicular transport, alter the synaptic plasticity (loss of synaptic integrity), and in the worst-case scenario maybe even result in loss of synaptic regulation leading to the accumulation of metabolic products in the synapse as is commonly seen in PD (Fahn, 2010; Schapira and Jenner, 2011; Picconi et al., 2012).

Protein complexes at the postsynaptic site have been implicated in several PD animal models as a key factor in PD (Belluzzi et al., 2012; de la Fuente-Fernández et al., 2004; Roach, 2007). However, recent research has shown that the proteins linked to PD such as SNCA and LRRK2 play a critical role at the presynaptic site (Li et al., 2010; Esposito et al., 2012). These studies have demonstrated that PD progression is affected by the dopamine release into the synaptic cleft which relies on functional presynaptic vesicular transport; synaptic vesicle trafficking, and the maintained regulation of presynaptic plasticity. All of these presynaptic features proposed to contribute to parkinsonism in animal models are closely associated with the function of NRXN2α.

As stated before, the MAF of the variant in the local control population and in populations in online databases was 0 % for the *NRXN2α*. p.G849D mutation, meaning that it is not present in the general population. This result further supports the causative nature of the variant, because disease-causing mutations are not expected to be common in the general population due to negative selection acting against them (Wong et al., 2003). The variant was absent in our cohort of 671 PD patients of multiple ethnicities. Also, we screened Annex, a web-based database that contains over 3000 exomes of patients with neurological diseases (Table 3.6, Results), predominantly PD, from one of our collaborators (Prof. Matt Farrer, Canada). None of these patients had the p.G849D variant however, 23 other likely pathogenic variants in the *NRXN2* gene were shared amongst 50 patients diagnosed with PD. This includes 12 indels shared by 22 of the PD patients.

Furthermore, in order to determine the expression of endogenous NRXN2α, Western blotting was done using the protein lysates from an un-transfected SH-SY5Y human Neuroblastoma cell line and a commercially available polyclonal NRXN2α antibody. The Western blots exhibited no non-specific binding or background and only a single band was observed. The size of the band (270 kDa) however, did not correspond to the expected size of the endogenous protein (185 kDa in lysates from a mouse Neuroblastoma cell line N2A). A search of the Celera mouse and human genome database (https://www.ocf.berkeley.edu/~edy/genome/celera.html) revealed that the mouse and human homologs of the gene exhibit exon and intron structures relatively similar in size. The *NRXN2α Mus musculus* homolog was 103,6 kb and the *Homo sapiens* homolog was 106,4 kb (Tabuchi and Südhof, 2002). Therefore, the difference between the observed and expected sizes of the protein are not a result of species-specific gene structure.

Membrane proteins like NRXN2α which are associated with lipid bilayers, have been shown to be difficult to denature in aqueous WB solutions because of high hydrophobicity and poor solubility often forming bulky quaternary protein structures (Kubicek et al., 2014). This could explain the highly specific single protein band observed as unresolved quaternary NRXN2α proteins. The use of stronger detergents has been suggested to increase the purification and solubilisation (Mancia and Love, 2010). Another reason we might be seeing the larger NRXN2α molecule could be because it is bound to another protein such as NLNGs which are typically about 100 kDa in size.

4.5. Study Limitations

Our study had a number of limitations. Though useful, the functional prediction tools are based on assumptions, and all tools have various short-comings. The developers of these tools emphasise that the terms 'damaging' or 'deleterious' should never be logically equated with causal of disease phenotype.

The variant filtering strategy used follows what has previously been suggested in the literature. This influenced the type of variants filtered out at each stage and led to some variants making it to higher or lower ranking which may be incorrect. The filtration and interpretation of the data was also limited by the current knowledge about the genes and the information on them that is available in online databases. This incomplete knowledge about the function and disease associations of certain genes that may also lead to candidate variants being erroneously discarded.

Another limitation of our study is in the limited sequencing scope of WES. One disadvantage of this approach is its inability to accurately detect structural variants such as inversions and translocations, as well as other types of variants such as non-coding variants, deletions, copy number variations and repeat expansions (Botstein and Risch, 2003) (Krumm et al., 2012). Some non-coding variants may be important for controlling gene transcriptional regulation or splicing (Fromer et al., 2012). The target coverage along the length of exons in WES is usually uneven which causes errors in downstream analysis such as false-positive calls in variant calling analysis. Factors that may affect coverage in WES include the size of the exon, the presence or absence of repeat elements, GC content and segmental duplications (Belkadi et al., 2015; Kozarewa et al., 2009).

A technical limitation of this study is that we were unable to generate full-length *NRXN2α* cDNA for functional studies. cDNA of this gene is not currently commercially available. We had initially planned to generate this *NRXN2α* cDNA to clone the wild-type and mutant inserts into pDsRed-Monomer-C1 vectors. This was to be transfected into the SH-SY5Y Neuroblastoma cell line for further study of the wild-type and mutant form of the gene in a neuronal cell model. Unfortunately, we were unable to overcome the many technical challenges of generating full-length *NRXN2α* cDNA and consequently, only investigated the endogenous NRXN2α protein expression levels of the cell line using a commercially available antibody. This does lay a foundation for using this cell line to further study the NRXN2α protein.

For further functional study of the variant, tissue samples like skin and brain biopsy were not available due to these procedures being invasive and brain biopsies are only performed

post-mortem. Lastly, because the analysis process is not an automated one, there is the chance for human error in the experiments, and interpretation of the data.

4.6. Future Work

Future work should include the verification of the protein detected by the anti- NRXN2α antibody, this could be done by protein identification using mass spectrometry tools such as MALDI-TOF Mass Spectrometry and Q-TOF Mass Spectrometry. The endogenous protein levels of NRXN2α could be assessed in other human cell types using the same primary antibody, to determine if the large sized bands observed in the SH-SY5Y cells are cell specific. To attain full-length *NRXN2α*, gene synthesis services could be utilised to create the wild-type construct of the gene. Alternatively, *in vitro* CRISPR/Cas9 could also be used to produce the desired NRXN2α variant to create the mutant endogenous cell model for further functional investigation.

Ultimately, the functional assessment of the *NRXN2α* variant is necessary to establish its biological significance. Proteomic studies should be pursued to determine the potential effect of the variant on the structure, physiological role and function of the protein. In addition, the other three candidate genes *RFT1*, *TEP1* and *CFAP65* should also be functionally investigated for their possible contribution to the disease pathogenesis. As sequencing technology is rapidly developing with frequent improvements to the technology and its bioinformatics solutions, the previously acquired WES data should be re-analysed for new insights. In future, individual 11.902 who was considered to be a phenocopy could be considered a true PD patient and the data re-evaluated to determine if the initial analysis was fair and no genetic variants were mistakenly discarded. Additionally, ongoing follow up should be maintained with the family to recruit any newly affected individuals.

4.7. Conclusion

In PD research, genetic studies conducted on multiplex families have led to the identification of genes that have elucidated some of the pathways involved in disease pathogenesis. This includes *LRRK2*, *Parkin, PINK1 DJ-1* and *SNCA* (Nuytemans et al., 2010). With the advancement of sequencing technologies, it is anticipated that additional PD causative genes will be identified. The present study provides several lines of evidence implicating a mutation in *NRXN2α* as a likely novel genetic variant that could contribute to PD pathology. To our knowledge, this is the first report of an association between *NRXN2α* and PD. As a candidate *NRXN2α* is well-suited for future efforts directed at in-depth functional dissection. Elucidating the pathologic potential of variants and their

relative contribution to the phenotype may provide critical insight into disease mechanisms and potentially lead to development of improved therapeutic modalities for this debilitating disorder.

Reference List

Abdel-Salam, O.M.E., 2019. Use of Herbal Products/Alternative Medicines in Neurodegenerative Diseases (Alzheimer's Disease and Parkinson's Disease), in: Singh, S., Joshi, N. (Eds.), Pathology, Prevention and Therapeutics of Neurodegenerative Disease. Springer Singapore, Singapore, pp. 279–301. https://doi.org/10.1007/978-981-13-0944-1_24

Abou-Sleiman, P.M., Muqit, M.M.K., McDonald, N.Q., Yang, Y.X., Gandhi, S., Healy, D.G., Harvey, K., Harvey, R.J., Deas, E., Bhatia, K., Quinn, N., Lees, A., Latchman, D.S., Wood, N.W., 2006. A heterozygous effect for PINK1 mutations in Parkinson's disease? Ann. Neurol. 60, 414–419. https://doi.org/10.1002/ana.20960

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc Hum Genet 0 7, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Ahlskog, J.E., 2009. Parkin and PINK1 parkinsonism may represent nigral mitochondrial cytopathies distinct from Lewy body Parkinson's disease. Parkinsonism Relat Disord 15, 721–727. https://doi.org/10.1016/j.parkreldis.2009.09.010

Akinyemi, R.O., 2012. Epidemiology of Parkinsonism and Parkinson's disease in Sub-Saharan Africa: Nigerian profile. J Neurosci Rural Pract 3, 233–234. https://doi.org/10.4103/0976-3147.102586

Alfred Goldscheider, 1898. Physiologie des Muskelsinnes. Leipzig 432.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Andrews, S., 2016. FASTQC [http://www. bioinformatics. babraham. ac. uk/projects/fastqc/]. Accessed.

Antonini, A., Bauer, L., Dohin, E., Oertel, W.H., Rascol, O., Reichmann, H., Schmid, M., Singh, P., Tolosa, E., Chaudhuri, K.R., 2015. Effects of rotigotine transdermal patch in patients with Parkinson's disease presenting with non-motor symptoms – results of a double-blind, randomized, placebo-controlled trial. Eur J Neurol 22, 1400–1407. https://doi.org/10.1111/ene.12757

Appel-Cresswell, S., Rajput, A.H., Sossi, V., Thompson, C., Silva, V., McKenzie, J., Dinelle, K., McCormick, S.E., Vilariño-Güell, C., Stoessl, A.J., Dickson, D.W., Robinson, C.A.,

Farrer, M.J., Rajput, A., 2014. Clinical, positron emission tomography, and pathological studies of DNAJC13 p.N855S Parkinsonism. Mov. Disord. 29, 1684–1687. https://doi.org/10.1002/mds.26019

Asakawa, S., Tsunematsu, K., Takayanagi, A., Sasaki, T., Shimizu, A., Shintani, A., Kawasaki, K., Mungall, A.J., Beck, S., Minoshima, S., Shimizu, N., 2001. The Genomic Structure and Promoter Region of the Human Parkin Gene. Biochemical and Biophysical Research Communications 286, 863–868. https://doi.org/10.1006/bbrc.2001.5490

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., Shendure, J., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. 12, 745–755. https://doi.org/10.1038/nrg3031

Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W.A., Jiang, H., Feng, G., 2014. Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. Cancer Inform 13, 67–82. https://doi.org/10.4137/CIN.S13779

Bardien, S., Keyser, R., Lombard, D., Du Plessis, M., Human, H., Carr, J., 2010. Novel non-sense GCH1 mutation in a South African family diagnosed with dopa-responsive dystonia. European Journal of Neurology 17, 510–512. https://doi.org/10.1111/j.1468-1331.2009.02725.x

Bardien, S., Keyser, R., Yako, Y., Lombard, D., Carr, J., 2009. Molecular analysis of the parkin gene in South African patients diagnosed with Parkinson's disease. Parkinsonism & Related Disorders 15, 116–121. https://doi.org/10.1016/j.parkreldis.2008.04.005

Bardien, Soraya, Marsberg, A., Keyser, R., Lombard, D., Lesage, S., Brice, A., Carr, J., 2010. LRRK2 G2019S mutation: frequency and haplotype data in South African Parkinson's disease patients. J Neural Transm 117, 847–853. https://doi.org/10.1007/s00702-010-0423-6

Barkhuizen, M., Anderson, D.G., van der Westhuizen, F.H., Grobler, A.F., 2017. A molecular analysis of the GBA gene in Caucasian South Africans with Parkinson's disease. Mol Genet Genomic Med 5, 147–156. https://doi.org/10.1002/mgg3.267

Baudhuin, L.M., Lagerstedt, S.A., Klee, E.W., Fadra, N., Oglesbee, D., Ferber, M.J., 2015. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. The Journal of Molecular Diagnostics 17, 456–461. https://doi.org/10.1016/j.jmoldx.2015.03.004

Baumann, R.J., Jameson, H.D., McKean, H.E., Haack, D.G., Weisberg, L.M., 1980. Cigarette smoking and Parkinson disease. Neurology 30, 839. https://doi.org/10.1212/WNL.30.8.839

Behnke, S., Berg, D., Naumann, M., Becker, G., 2005. Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound. Journal of Neurology, Neurosurgery & Psychiatry 76, 423–425. https://doi.org/10.1136/jnnp.2004.049221

Belin, A.C., Westerlund, M., 2008. Parkinson's disease: A genetic perspective. FEBS Journal 275, 1377–1383. https://doi.org/10.1111/j.1742-4658.2008.06301.x

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., Antipenko, A., Shang, L., Boisson, B., Casanova, J.-L., Abel, L., 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc. Natl. Acad. Sci. U.S.A. 112, 5473–5478. https://doi.org/10.1073/pnas.1418631112

Belluzzi, E., Greggio, E., Piccoli, G., 2012. Presynaptic dysfunction in Parkinson's disease: a focus on LRRK2. Biochem. Soc. Trans. 40, 1111–1116. https://doi.org/10.1042/BST20120124

Benabid, A.L., Chabardes, S., Mitrofanis, J., Pollak, P., 2009. Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. The Lancet Neurology 8, 67–81. https://doi.org/10.1016/S1474-4422(08)70291-6

Benamer, H.T.S., de Silva, R., Siddiqui, K.A., Grosset, D.G., 2008. Parkinson's disease in Arabs: A systematic review. Mov. Disord. 23, 1205–1210. https://doi.org/10.1002/mds.22041

Blanckenberg, J., Bardien, S., Glanzmann, B., Okubadejo, N.U., Carr, J.A., 2013. The prevalence and genetics of Parkinson's disease in sub-Saharan Africans. Journal of the Neurological Sciences 335, 22–25. https://doi.org/10.1016/j.jns.2013.09.010

Blanckenberg, J., Ntsapi, C., Carr, J.A., Bardien, S., 2014. EIF4G1 R1205H and VPS35 D620N mutations are rare in Parkinson's disease from South Africa. Neurobiology of Aging 35, 445.e1-445.e3. https://doi.org/10.1016/j.neurobiolaging.2013.08.023

Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C.J., Squitieri, F., Ibanez, P., Joosse, M., van Dongen, J.W., Vanacore, N., van Swieten, J.C., Brice, A., Meco, G., van Duijn, C.M., Oostra, B.A., Heutink, P., 2003. Mutations in

the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299, 256–259. https://doi.org/10.1126/science.1077209

Bonifati, V., Rohé, C.F., Breedveld, G.J., Fabrizio, E., De Mari, M., Tassorelli, C., Tavella, A., Marconi, R., Nicholl, D.J., Chien, H.F., Fincati, E., Abbruzzese, G., Marini, P., De Gaetano, A., Horstink, M.W., Maat-Kievit, J.A., Sampaio, C., Antonini, A., Stocchi, F., Montagna, P., Toni, V., Guidi, M., Dalla Libera, A., Tinazzi, M., De Pandis, F., Fabbrini, G., Goldwurm, S., de Klein, A., Barbosa, E., Lopiano, L., Martignoni, E., Lamberti, P., Vanacore, N., Meco, G., Oostra, B.A., Italian Parkinson Genetics Network, 2005. Early-onset parkinsonism associated with PINK1 mutations: frequency, genotypes, and phenotypes. Neurology 65, 87–95. https://doi.org/10.1212/01.wnl.0000167546.39375.82

Booker, D., 2012. Parkinson Disease and Rehabilitation Administration and Services. Southern Illinois University Carbondale.

Bordoli, L., Schwede, T., 2012. Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal. Methods Mol Biol 857, 107–136. https://doi.org/10.1007/978-1-61779-588-6_5

Botstein, D., Risch, N., 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl, 228–237. https://doi.org/10.1038/ng1090

Bower, J.H., Maraganore, D.M., McDonnell, S.K., Rocca, W.A., 1999. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. Neurology 52, 1214–1220.

Braak, H., Del Tredici, K., Bratzke, H., Hamm-Clement, J., Sandmann-Keil, D., Rüb, U., 2002. Staging of the intracerebral inclusion body pathology associated with idiopathic Parkinson's disease (preclinical and clinical stages). J. Neurol. 249 Suppl 3, III/1-5.

Braak, H., Sastre, M., Bohl, J.R.E., Vos, R.A.I. de, Tredici, K.D., 2007. Parkinson's disease: lesions in dorsal horn layer I, involvement of parasympathetic and sympathetic pre- and postganglionic neurons. Acta Neuropathol 113, 421–429. https://doi.org/10.1007/s00401-007-0193-x

Braak, H., Tredici, K.D., Rüb, U., de Vos, R.A.I., Jansen Steur, E.N.H., Braak, E., 2003. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiology of Aging 24, 197–211. https://doi.org/10.1016/S0197-4580(02)00065-9

Brink, P.A., Crotti, L., Corfield, V., Goosen, A., Durrheim, G., Hedley, P., Heradien, M., Geldenhuys, G., Vanoli, E., Bacchini, S., Spazzolini, C., Lundquist, A.L., Roden, D.M., George, A.L., Schwartz, P.J., 2005. Phenotypic Variability and Unusual Clinical Severity of Congenital Long-QT Syndrome in a Founder Population. Circulation 112, 2602–2610. https://doi.org/10.1161/CIRCULATIONAHA.105.572453

Burke, R.E., Dauer, W.T., Vonsattel, J.P.G., 2008. A Critical Evaluation of The Braak Staging Scheme for Parkinson's Disease. Ann Neurol 64, 485–491. https://doi.org/10.1002/ana.21541

Cacabelos, R., 2017a. Parkinson's Disease: From Pathogenesis to Pharmacogenomics. International Journal of Molecular Sciences 18, 551. https://doi.org/10.3390/ijms18030551

Cacabelos, R., 2017b. Parkinson's Disease: From Pathogenesis to Pharmacogenomics. Int J Mol Sci 18. https://doi.org/10.3390/ijms18030551

Calne, D.B., Langston, J.W., 1983. Aetiology of Parkinson's disease. Lancet 2, 1457–1459.

Campbell, M.C., Tishkoff, S.A., 2008. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. Annual Review of Genomics and Human Genetics 9, 403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Campenhausen, S. von, Bornschein, B., Wick, R., Bötzel, K., Sampaio, C., Poewe, W., Oertel, W., Siebert, U., Berger, K., Dodel, R., 2005. Prevalence and incidence of Parkinson's disease in Europe. European Neuropsychopharmacology 15, 473–490. https://doi.org/10.1016/j.euroneuro.2005.04.007

Carr, J., Guella, I., Szu-Tu, C., Boolay, S., Glanzmann, B., Farrer, M.J., Bardien, S., 2016. Double homozygous mutations (R275W and M432V) in the ParkinGene associated with late-onset Parkinson's disease: Homozygous Mutations in Parkin Gene. Movement Disorders 31, 423–425. https://doi.org/10.1002/mds.26524

Challis, D., Yu, J., Evani, U.S., Jackson, A.R., Paithankar, S., Coarfa, C., Milosavljevic, A., Gibbs, R.A., Yu, F., 2012. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 13, 8. https://doi.org/10.1186/1471-2105-13-8

Chartier-Harlin, M.-C., Dachsel, J.C., Vilariño-Güell, C., Lincoln, S.J., Leprêtre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.-S., Le Rhun, E., Mutez, E., Larvor, L., Duflot, A., Vanbesien-Mailliot, C., Kreisler, A., Ross, O.A., Nishioka,

K., Soto-Ortolaza, A.I., Cobb, S.A., Melrose, H.L., Behrouz, B., Keeling, B.H., Bacon, J.A., Hentati, E., Williams, L., Yanagiya, A., Sonenberg, N., Lockhart, P.J., Zubair, A.C., Uitti, R.J., Aasly, J.O., Krygowska-Wajs, A., Opala, G., Wszolek, Z.K., Frigerio, R., Maraganore, D.M., Gosal, D., Lynch, T., Hutchinson, M., Bentivoglio, A.R., Valente, E.M., Nichols, W.C., Pankratz, N., Foroud, T., Gibson, R.A., Hentati, F., Dickson, D.W., Destée, A., Farrer, M.J., 2011. Translation Initiator EIF4G1 Mutations in Familial Parkinson Disease. Am J Hum Genet 89, 398–406. https://doi.org/10.1016/j.ajhg.2011.08.009

Chen, L.Y., Jiang, M., Zhang, B., Gokce, O., Südhof, T.C., 2017. Conditional Deletion of All Neurexins Defines Diversity of Essential Synaptic Organizer Functions for Neurexins. Neuron 94, 611-625.e4. https://doi.org/10.1016/j.neuron.2017.04.011

Cheng, H.-C., Ulane, C.M., Burke, R.E., 2010. Clinical Progression in Parkinson's Disease and the Neurobiology of Axons. Ann Neurol 67, 715–725. https://doi.org/10.1002/ana.21995

Cherukuri, P.F., Maduro, V., Fuentes-Fajardo, K.V., Lam, K., Adams, D.R., Tifft, C.J., Mullikin, J.C., Gahl, W.A., Boerkoel, C.F., 2015. Replicate exome-sequencing in a multiple-generation family: improved interpretation of next-generation sequencing data. BMC Genomics 16. https://doi.org/10.1186/s12864-015-2107-y

Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Cilia, R., Sironi, F., Akpalu, A., Cham, M., Sarfo, F.S., Brambilla, T., Bonetti, A., Amboni, M., Goldwurm, S., Pezzoli, G., 2011. Screening LRRK2 gene mutations in patients with Parkinson's disease in Ghana. J Neurol 259, 569–570. https://doi.org/10.1007/s00415-011-6210-y

Clarke, C.E., Patel, S., Ives, N., Rick, C.E., Woolley, R., Wheatley, K., Walker, M.F., Zhu, S., Kandiyali, R., Yao, G., Sackley, C.M., Group, on behalf of the P.R.C., 2016. UK Parkinson's Disease Society Brain Bank Diagnostic Criteria. NIHR Journals Library.

Colosimo, C., Morgante, L., Antonini, A., Barone, P., Avarello, T.P., Bottacchi, E., Cannas, A., Ceravolo, M.G., Ceravolo, R., Cicarelli, G., Gaglio, R.M., Giglia, L., Iemolo, F., Manfredi, M., Meco, G., Nicoletti, A., Pederzoli, M., Petrone, A., Pisani, A., Pontieri, F.E., Quatrale, R., Ramat, S., Scala, R., Volpe, G., Zappulla, S., Bentivoglio, A.R., Stocchi, F., Trianni, G., Dotto, P.D., Simoni, L., Marconi, R., Group, and P.S., 2010. Non-motor

symptoms in atypical and secondary parkinsonism: the PRIAMO study. J Neurol 257, 5. https://doi.org/10.1007/s00415-009-5255-7

Cosnett, J.E., 1973. Neurological disease in Natal. Tropical Neurology 270.

Craig, A.M., Kang, Y., 2007. Neurexin-neuroligin signaling in synapse development. Curr. Opin. Neurobiol. 17, 43–52. https://doi.org/10.1016/j.conb.2007.01.011

Cubo, E., Doumbe, J., Martinez-Martin, P., Rodriguez-Blazquez, C., Kuate, C., Mariscal, N., Lopez, I., Noubissi, G., Mapoure, Y.N., Jon, J.L., Mbahe, S., Tchaleu, B., Catalan, M.-J., 2014. Comparison of the clinical profile of Parkinson's disease between Spanish and Cameroonian Cohorts. Journal of the Neurological Sciences 336, 122–126. https://doi.org/10.1016/j.jns.2013.10.021

Dachtler, J., Glasper, J., Cohen, R.N., Ivorra, J.L., Swiffen, D.J., Jackson, A.J., Harte, M.K., Rodgers, R.J., Clapcote, S.J., 2014. Deletion of α-neurexin II results in autism-related behaviors in mice. Transl Psychiatry 4, e484. https://doi.org/10.1038/tp.2014.123

Dachtler, J., Ivorra, J.L., Rowland, T.E., Lever, C., Rodgers, R.J., Clapcote, S.J., 2015. Heterozygous Deletion of α-Neurexin I or α-Neurexin II Results in Behaviors Relevant to Autism and Schizophrenia. Behav Neurosci 129, 765–776. https://doi.org/10.1037/bne0000108

Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., Batzoglou, S., 2010. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLOS Computational Biology 6, e1001025. https://doi.org/10.1371/journal.pcbi.1001025

de la Fuente-Fernández, R., Schulzer, M., Mak, E., Calne, D.B., Stoessl, A.J., 2004. Presynaptic mechanisms of motor fluctuations in Parkinson's disease: a probabilistic model. Brain 127, 888–899. https://doi.org/10.1093/brain/awh102

de Lau, L.M., Breteler, M.M., 2006. Epidemiology of Parkinson's disease. The Lancet Neurology 5, 525–535.

de Rijk, M.C., Launer, L.J., Berger, K., Breteler, M.M., Dartigues, J.F., Baldereschi, M., Fratiglioni, L., Lobo, A., Martinez-Lage, J., Trenkwalder, C., Hofman, A., 2000. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 54, S21-23.

de Wit, J., Sylwestrak, E., O'Sullivan, M.L., Otto, S., Tiglio, K., Savas, J.N., Yates, J.R., Comoletti, D., Taylor, P., Ghosh, A., 2009. LRRTM2 Interacts with Neurexin1 and

Regulates Excitatory Synapse Formation. Neuron 64, 799–806. https://doi.org/10.1016/j.neuron.2009.12.019

Deng, H.-X., Shi, Y., Yang, Y., Ahmeti, K.B., Miller, N., Huang, C., Cheng, L., Zhai, H., Deng, S., Nuytemans, K., Corbett, N.J., Kim, M.J., Deng, H., Tang, B., Yang, Z., Xu, Y., Chan, P., Huang, B., Gao, X.-P., Song, Z., Liu, Z., Fecto, F., Siddique, N., Foroud, T., Jankovic, J., Ghetti, B., Nicholson, D.A., Krainc, D., Melen, O., Vance, J.M., Pericak-Vance, M.A., Ma, Y.-C., Rajput, A.H., Siddique, T., 2016. Identification of TMEM230 mutations in familial Parkinson's disease. Nat. Genet. 48, 733–739. https://doi.org/10.1038/ng.3589

Department of Economic and Social Affairs, 2016. Policy - United Nations Population Division [WWW Document]. United Nations Org. URL http://www.un.org/en/development/desa/population/theme/policy/index.shtml (accessed 5.3.16).

Di Fonzo, A., Dekker, M.C.J., Montagna, P., Baruzzi, A., Yonova, E.H., Correia Guedes, L., Szczerbinska, A., Zhao, T., Dubbel-Hulsman, L.O.M., Wouters, C.H., de Graaff, E., Oyen, W.J.G., Simons, E.J., Breedveld, G.J., Oostra, B.A., Horstink, M.W., Bonifati, V., 2009. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology 72, 240–245. https://doi.org/10.1212/01.wnl.0000338144.10967.2b

Di Monte, D.A., Lavasani, M., Manning-Bog, A.B., 2002a. Environmental factors in Parkinson's disease. Neurotoxicology 23, 487–502.

Di Monte, D.A., Lavasani, M., Manning-Bog, A.B., 2002b. Environmental Factors in Parkinson's Disease. NeuroToxicology 23, 487–502. https://doi.org/10.1016/S0161-813X(02)00099-2

Dick, F.D., De Palma, G., Ahmadi, A., Scott, N.W., Prescott, G.J., Bennett, J., Semple, S., Dick, S., Counsell, C., Mozzoni, P., Haites, N., Wettinger, S.B., Mutti, A., Otelea, M., Seaton, A., Söderkvist, P., Felice, A., 2007. Environmental risk factors for Parkinson's disease and parkinsonism: the Geoparkinson study. Occup Environ Med 64, 666–672. https://doi.org/10.1136/oem.2006.027003

Dorsey, E.R., Constantinescu, R., Thompson, J.P., Biglan, K.M., Holloway, R.G., Kieburtz, K., Marshall, F.J., Ravina, B.M., Schifitto, G., Siderowf, A., Tanner, C.M., 2007. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 68, 384–386. https://doi.org/10.1212/01.wnl.0000247740.47667.03

Dotchin, C., Msuya, O., Kissima, J., Massawe, J., Mhina, A., Moshy, A., Aris, E., Jusabani, A., Whiting, D., Masuki, G., Walker, R., 2008. The prevalence of Parkinson's disease in rural Tanzania. Mov. Disord. 23, 1567–1672. https://doi.org/10.1002/mds.21898

Driver, J.A., Logroscino, G., Gaziano, J.M., Kurth, T., 2009. Incidence and remaining lifetime risk of Parkinson disease in advanced age. Neurology 72, 432–438. https://doi.org/10.1212/01.wnl.0000341769.50075.bb

Drouet, V., Lesage, S., 2014. Synaptojanin 1 Mutation in Parkinson's Disease Brings Further Insight into the Neuropathological Mechanisms. Biomed Res Int 2014. https://doi.org/10.1155/2014/289728

Esposito, G., Ana Clara, F., Verstreken, P., 2012. Synaptic vesicle trafficking and Parkinson's disease. Dev Neurobiol 72, 134–144. https://doi.org/10.1002/dneu.20916

Fahn, S., 2010. Parkinson's disease: 10 years of progress, 1997–2007. Mov. Disord. 25, S2–S14. https://doi.org/10.1002/mds.22796

Farlow, J.L., Robak, L.A., Hetrick, K., Bowling, K., Boerwinkle, E., Coban-Akdemir, Z.H., Gambin, T., Gibbs, R.A., Gu, S., Jain, P., Jankovic, J., Jhangiani, S., Kaw, K., Lai, D., Lin, H., Ling, H., Liu, Y., Lupski, J.R., Muzny, D., Porter, P., Pugh, E., White, J., Doheny, K., Myers, R.M., Shulman, J.M., Foroud, T., 2016a. Whole-Exome Sequencing in Familial Parkinson Disease. JAMA Neurol 73, 68–75. https://doi.org/10.1001/jamaneurol.2015.3266

Farlow, J.L., Robak, L.A., Hetrick, K., Bowling, K., Boerwinkle, E., Coban-Akdemir, Z.H., Gambin, T., Gibbs, R.A., Gu, S., Jain, P., Jankovic, J., Jhangiani, S., Kaw, K., Lai, D., Lin, H., Ling, H., Liu, Y., Lupski, J.R., Muzny, D., Porter, P., Pugh, E., White, J., Doheny, K., Myers, R.M., Shulman, J.M., Foroud, T., 2016b. Whole-Exome Sequencing in Familial Parkinson Disease. JAMA Neurol 73, 68–75. https://doi.org/10.1001/jamaneurol.2015.3266

Femi, O., Ibrahim, A., Aliyu, S., 2012. Clinical profile of parkinsonian disorders in the tropics: Experience at Kano, northwestern Nigeria. Journal of Neurosciences in Rural Practice 3, 237. https://doi.org/10.4103/0976-3147.102589

Flanagan, S.E., Patch, A.-M., Ellard, S., 2010. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. Genetic Testing and Molecular Biomarkers 14, 533–537. https://doi.org/10.1089/gtmb.2010.0036

Follett, K.A., Weaver, F.M., Stern, M., Hur, K., Harris, C.L., Luo, P., Marks, W.J.J., Rothlind, J., Sagher, O., Moy, C., Pahwa, R., Burchiel, K., Hogarth, P., Lai, E.C., Duda, J.E., Holloway, K., Samii, A., Horn, S., Bronstein, J.M., Stoner, G., Starr, P.A., Simpson, R., Baltuch, G., De Salles, A., Huang, G.D., Reda, D.J., 2010. Pallidal versus Subthalamic Deep-Brain Stimulation for Parkinson's Disease [WWW Document]. http://dx.doi.org/10.1056/NEJMoa0907083. https://doi.org/10.1056/NEJMoa0907083

Forno, L.S., 1996. Neuropathology of Parkinson's Disease. J Neuropathol Exp Neurol 55, 259–272. https://doi.org/10.1097/00005072-199603000-00001

Frigerio, R., Fujishiro, H., Ahn, T.-B., Josephs, K.A., Maraganore, D.M., DelleDonne, A., Parisi, J.E., Klos, K.J., Boeve, B.F., Dickson, D.W., Ahlskog, E.J., 2011. Incidental Lewy Body Disease: Do some cases represent a preclinical stage of Dementia with Lewy Bodies? Neurobiol Aging 32, 857–863. https://doi.org/10.1016/j.neurobiolaging.2009.05.019

Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., Kirov, G., Sullivan, P.F., Hultman, C.M., Sklar, P., Purcell, S.M., 2012. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. 91, 597–607. https://doi.org/10.1016/j.ajhg.2012.08.005

Fuccillo, M.V., Földy, C., Gökce, Ö., Rothwell, P.E., Sun, G.L., Malenka, R.C., Südhof, T.C., 2015. Single-Cell mRNA Profiling Reveals Cell-Type-Specific Expression of Neurexin Isoforms. Neuron 87, 326–340. https://doi.org/10.1016/j.neuron.2015.06.028

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., Tomiyama, H., Nishioka, K., Hasegawa, K., Saiki, H., Satake, W., Mogushi, K., Sasaki, R., Kokubo, Y., Kuzuhara, S., Toda, T., Mizuno, Y., Uchiyama, Y., Ohno, K., Hattori, N., 2015a. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. The Lancet Neurology 14, 274–282. https://doi.org/10.1016/S1474-4422(14)70266-2

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., Tomiyama, H., Nishioka, K., Hasegawa, K., Saiki, H., Satake, W., Mogushi, K., Sasaki, R., Kokubo, Y., Kuzuhara, S., Toda, T., Mizuno, Y., Uchiyama, Y., Ohno, K., Hattori, N., 2015b. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. The Lancet Neurology 14, 274–282. https://doi.org/10.1016/S1474-4422(14)70266-2

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., Tomiyama, H., Nishioka, K., Hasegawa, K., Saiki, H., Satake, W., Mogushi, K., Sasaki, R., Kokubo, Y., Kuzuhara, S., Toda, T., Mizuno, Y., Uchiyama, Y., Ohno, K., Hattori, N., 2015c. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. The Lancet Neurology 14, 274–282. https://doi.org/10.1016/S1474-4422(14)70266-2

Futai, K., Hayashi, Y., 2008. Transsynaptic Regulation of Presynaptic Release Machinery in Central Synapses by Cell Adhesion Molecules, in: Molecular Mechanisms of Neurotransmitter Release, Contemporary Neuroscience. Humana Press, Totowa, NJ, pp. 315–326. https://doi.org/10.1007/978-1-59745-481-0_15

Gagliardi, M., Iannello, G., Colica, C., Annesi, G., Quattrone, A., 2017. Analysis of CHCHD2 gene in familial Parkinson's disease from Calabria. Neurobiol. Aging 50, 169.e5-169.e6. https://doi.org/10.1016/j.neurobiolaging.2016.10.022

Geldenhuys, G., Glanzmann, B., Lombard, D., Boolay, S., Carr, J., Bardien, S., 2014. Identification of a common founder couple for 40 South African Afrikaner families with Parkinson's disease. SAMJ: South African Medical Journal 104, 413–419.

Ghosh, R., Oak, N., Plon, S.E., 2017. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biology 18, 225. https://doi.org/10.1186/s13059-017-1353-5

Gibb, W.R., Lees, A.J., 1988. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry 51, 745–752.

Gilks, W.P., Abou-Sleiman, P.M., Gandhi, S., Jain, S., Singleton, A., Lees, A.J., Shaw, K., Bhatia, K.P., Bonifati, V., Quinn, N.P., Lynch, J., Healy, D.G., Holton, J.L., Revesz, T., Wood, N.W., 2005. A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet 365, 415–416. https://doi.org/10.1016/S0140-6736(05)17830-1

Giri, A., Mok, K.Y., Jansen, I., Sharma, M., Tesson, C., Mangone, G., Lesage, S., Bras, J.M., Shulman, J.M., Sheerin, U.-M., International Parkinson's Disease Consortium (IPDGC), Díez-Fairen, M., Pastor, P., Martí, M.J., Ezquerra, M., Tolosa, E., Correia-Guedes, L., Ferreira, J., Amin, N., van Duijn, C.M., van Rooij, J., Uitterlinden, A.G., Kraaij, R., Nalls, M., Simón-Sánchez, J., 2017. Lack of evidence for a role of genetic variation in TMEM230 in the risk for Parkinson's disease in the Caucasian population. Neurobiol. Aging 50, 167.e11-167.e13. https://doi.org/10.1016/j.neurobiolaging.2016.10.004

Glanzmann, B., 2013. An investigation into the molecular aetiology of Parkinson's disease in South African patients. Stellenbosch University, Stellenbosch.

Glanzmann, B., Lombard, D., Carr, J., Bardien, S., 2013. Screening of two indel polymorphisms in the 5′UTR of the DJ-1 gene in South African Parkinson's disease patients. J Neural Transm 121, 135–138. https://doi.org/10.1007/s00702-013-1094-x

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. Nat Biotechnol 27, 182–189. https://doi.org/10.1038/nbt.1523

Golbe, L.I., Lazzarini, A.M., Spychala, J.R., Johnson, W.G., Stenroos, E.S., Mark, M.H., Sage, J.I., 2001. The tau A0 allele in Parkinson's disease. Movement Disorders 16, 442–447. https://doi.org/10.1002/mds.1087

Gordon, A., Hannon, G.J., 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) http://hannonlab. cshl. edu/fastx_toolkit 5.

Gorell, J.M., Rybicki, B.A., Johnson, C.C., Peterson, E.L., 1999. Smoking and Parkinson's disease. Neurology 52, 115. https://doi.org/10.1212/WNL.52.1.115

Grandinetti, A., Morens, D.M., Reed, D., MacEachern, D., 1994. Prospective Study of Cigarette Smoking and the Risk of Developing Idiopathic Parkinson's Disease. Am J Epidemiol 139, 1129–1138. https://doi.org/10.1093/oxfordjournals.aje.a116960

Greeff, J.M., 2007. Deconstructing Jaco: genetic heritage of an Afrikaner. Ann. Hum. Genet. 71, 674–688. https://doi.org/10.1111/j.1469-1809.2007.00363.x

Groiss, S.J., Wojtecki, L., Südmeyer, M., Schnitzler, A., 2009. Deep Brain Stimulation in Parkinson's Disease. Ther Adv Neurol Disord 2, 20–28. https://doi.org/10.1177/1756285609339382

Group, T.P.S., 2006. Randomized placebo-controlled study of the nicotinic agonist SIB-1508Y in Parkinson disease. Neurology 66, 408–410. https://doi.org/10.1212/01.wnl.0000196466.99381.5c

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723. https://doi.org/10.1002/elps.1150181505

Haaxma, C.A., Bloem, B.R., Borm, G.F., Oyen, W.J.G., Leenders, K.L., Eshuis, S., Booij, J., Dluzen, D.E., Horstink, M.W.I.M., 2007. Gender differences in Parkinson's disease. J Neurol Neurosurg Psychiatry 78, 819–824. https://doi.org/10.1136/jnnp.2006.103788

Haeuptle, M.A., Pujol, F.M., Neupert, C., Winchester, B., Kastaniotis, A.J., Aebi, M., Hennet, T., 2008. Human RFT1 Deficiency Leads to a Disorder of N-Linked Glycosylation. The American Journal of Human Genetics 82, 600–606. https://doi.org/10.1016/j.ajhg.2007.12.021

Haimanot, R.T.-, Abebe, M., Gebre-Mariam, A., Forsgren, L., Heijbel, J., Holmgren, G., Ekstedt, J., 1990. Community-Based Study of Neurological Disorders in Rural Central Ethiopia. NED 9, 263–277. https://doi.org/10.1159/000110783

HALL, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, in: Nucleic Acids Symp. Ser. pp. 95–98.

Halliday, G., Lees, A., Stern, M., 2011. Milestones in Parkinson's disease--clinical and pathologic features. Mov. Disord. 26, 1015–1021. https://doi.org/10.1002/mds.23669

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, I., Robinson, M.O., 1997. A mammalian telomerase-associated protein. Science 275, 973–977.

Hauser, R.A., Pahwa, R., McClain, T.A., Lyons, K.E., 2016. Parkinson Disease: Practice Essentials, Background, Anatomy.

Hayden, M.R., Hopkins, H.C., Macrea, M., Beighton, P.H., 1980. The origin of Huntington's chorea in the Afrikaner population of South Africa. S Afr Med J 58, 197–200.

Haylett, W.L., Keyser, R.J., du Plessis, M.C., van der Merwe, C., Blanckenberg, J., Lombard, D., Carr, J., Bardien, S., 2012. Mutations in the parkin gene are a minor cause of Parkinson's disease in the South African population. Parkinsonism & Related Disorders 18, 89–92. https://doi.org/10.1016/j.parkreldis.2011.09.022

Heese, J.A., 1971. Die herkoms van die Afrikaner, 1657-1867.

Helenius, J., Ng, D.T.W., Marolda, C.L., Walter, P., Valvano, M.A., Aebi, M., 2002. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. Nature 415, 447–450. https://doi.org/10.1038/415447a

Hellenbrand, W., Seidler, A., Robra, B.P., Vieregge, P., Oertel, W.H., Joerg, J., Nischan, P., Schneider, E., Ulm, G., 1997. Smoking and Parkinson's disease: a case-control study in Germany. Int J Epidemiol 26, 328–339. https://doi.org/10.1093/ije/26.2.328

Hertzman, C., Wiens, M., Bowering, D., Snow, B., Calne, D., 1990. Parkinson's disease: A case-control study of occupational and environmental risk factors. American Journal of Industrial Medicine 17, 349–355. https://doi.org/10.1002/ajim.4700170307

Hirsch, L., Jette, N., Frolkis, A., Steeves, T., Pringsheim, T., 2016. The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. NED 46, 292–300. https://doi.org/10.1159/000445751

Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J., 1992. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55, 181–184. https://doi.org/10.1136/jnnp.55.3.181

Iacono, D., Geraci-Erck, M., Rabin, M.L., Adler, C.H., Serrano, G., Beach, T.G., Kurlan, R., 2015. Parkinson disease and incidental Lewy body disease. Neurology 85, 1670–1679. https://doi.org/10.1212/WNL.0000000000002102

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79, 368–376. https://doi.org/10.1136/jnnp.2007.131045

Jansen, I.E., Bras, J.M., Lesage, S., Schulte, C., Gibbs, J.R., Nalls, M.A., Brice, A., Wood, N.W., Morris, H., Hardy, J.A., Singleton, A.B., Gasser, T., Heutink, P., Sharma, M., IPDGC, 2015. CHCHD2 and Parkinson's disease. Lancet Neurol 14, 678–679. https://doi.org/10.1016/S1474-4422(15)00094-0

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Kapur, S.S., Stebbins, G.T., Goetz, C.G., 2012. Vibration Therapy for Parkinson's Disease: Charcot's Studies Revisited. Journal of Parkinson's Disease 2, 23–27. https://doi.org/10.3233/JPD-2012-12079

Karayiorgou, M., Torrington, M., Abecasis, G.R., Pretorius, H., Robertson, B., Kaliski, S., Lay, S., Sobin, C., Möller, N., Lundy, S.L., Blundell, M.L., Gogos, J.A., Roos, J.L., 2004. Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. Am. J. Med. Genet. 124B, 20–28. https://doi.org/10.1002/ajmg.b.20090

Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E., Cummings, B.B., Birnbaum, D., Daly, M.J.,

MacArthur, D.G., 2017. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res 45, D840–D845. https://doi.org/10.1093/nar/gkw971

Keyser, R.J., Lesage, S., Brice, A., Carr, J., Bardien, S., 2010a. Assessing the prevalence of PINK1 genetic variants in South African patients diagnosed with early- and late-onset Parkinson's disease. Biochemical and Biophysical Research Communications 398, 125–129. https://doi.org/10.1016/j.bbrc.2010.06.049

Keyser, R.J., Lesage, S., Brice, A., Carr, J., Bardien, S., 2010b. Assessing the prevalence of PINK1 genetic variants in South African patients diagnosed with early- and late-onset Parkinson's disease. Biochemical and Biophysical Research Communications 398, 125–129. https://doi.org/10.1016/j.bbrc.2010.06.049

Keyser, R.J., Lombard, D., Veikondis, R., Carr, J., Bardien, S., 2009a. Analysis of exon dosage using MLPA in South African Parkinson's disease patients. Neurogenetics 11, 305–312. https://doi.org/10.1007/s10048-009-0229-6

Keyser, R.J., Van Der Merwe, L., Venter, M., Kinnear, C., Warnich, L., Carr, J., Bardien, S., 2009b. Identification of a novel functional deletion variant in the 5'-UTR of the DJ-1 gene. https://doi.org/10.1186/1471-2350-10-105

Khodadadi, H., Azcona, L.J., Aghamollaii, V., Omrani, M.D., Garshasbi, M., Taghavi, S., Tafakhori, A., Shahidi, G.A., Jamshidi, J., Darvish, H., Paisán-Ruiz, C., 2017. PTRHD1 (C2orf79) mutations lead to autosomal recessive intellectual disability and parkinsonism. Mov Disord 32, 287–291. https://doi.org/10.1002/mds.26824

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315. https://doi.org/10.1038/ng.2892

Kisoli, A., Gray, W.K., Dotchin, C.L., Orega, G., Dewhurst, F., Paddick, S.-M., Longdon, A., Chaote, P., Dewhurst, M., Walker, R.W., 2015. Levels of functional disability in elderly people in Tanzania with dementia, stroke and Parkinson's disease. Acta Neuropsychiatrica 27, 206–212. https://doi.org/10.1017/neu.2015.9

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., Shimizu, N., 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392, 605–608. https://doi.org/10.1038/33416

Klein, C., Chuang, R., Marras, C., Lang, A.E., 2011. The curious case of phenocopies in families with genetic Parkinson's disease. Mov. Disord. 26, 1793–1802. https://doi.org/10.1002/mds.23853

Klein, C., Westenberger, A., 2012. Genetics of Parkinson's Disease. Cold Spring Harb Perspect Med 2. https://doi.org/10.1101/cshperspect.a008888

Knoll, D.P., de Vries, W.N., De Wet, W.J., 1988. Genealogie van'n familie met osteogenesis imperfecta: Die stamregister van Pieter Willem Adriaan Piek. Familia 25, 46–64.

Ko, J., Fuccillo, M.V., Malenka, R.C., Südhof, T.C., 2009. LRRTM2 Functions as a Neurexin Ligand in Promoting Excitatory Synapse Formation. Neuron 64, 791–798. https://doi.org/10.1016/j.neuron.2009.12.012

Kohanski, M.A., Dwyer, D.J., Collins, J.J., 2010. How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol 8, 423–435. https://doi.org/10.1038/nrmicro2333

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Koschmidder, E., Weissbach, A., Brüggemann, N., Kasten, M., Klein, C., Lohmann, K., 2016. A nonsense mutation in CHCHD2 in a patient with Parkinson disease. Neurology 86, 577. https://doi.org/10.1212/WNL.0000000000002361

Kowal, S.L., Dall, T.M., Chakrabarti, R., Storm, M.V., Jain, A., 2013. The current and projected economic burden of Parkinson's disease in the United States. Mov. Disord. 28, 311–318. https://doi.org/10.1002/mds.25292

Kozarewa, I., Ning, Z., Quail, M.A., Sanders, M.J., Berriman, M., Turner, D.J., 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods 6, 291–295. https://doi.org/10.1038/nmeth.1311

Krebs, C.E., Karkheiran, S., Powell, J.C., Cao, M., Makarov, V., Darvish, H., Di Paolo, G., Walker, R.H., Shahidi, G.A., Buxbaum, J.D., De Camilli, P., Yue, Z., Paisán-Ruiz, C., 2013. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Hum. Mutat. 34, 1200–1207. https://doi.org/10.1002/humu.22372

Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., NHLBI Exome Sequencing Project, Quinlan, A.R., Nickerson, D.A., Eichler, E.E., 2012. Copy number

variation detection and genotyping from exome sequence data. Genome Res. 22, 1525–1532. https://doi.org/10.1101/gr.138115.112

Kubicek, J., Block, H., Maertens, B., Spriestersbach, A., Labahn, J., 2014. Expression and purification of membrane proteins. Meth. Enzymol. 541, 117–140. https://doi.org/10.1016/B978-0-12-420119-4.00010-0

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4, 1073–1081. https://doi.org/10.1038/nprot.2009.86

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R.,

McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Langston, J.W., 2006. The parkinson's complex: Parkinsonism is just the tip of the iceberg. Ann Neurol. 59, 591–596. https://doi.org/10.1002/ana.20834

Langston, J.W., Ballard, P., Tetrud, J.W., Irwin, I., 1983. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science 219, 979–980.

Lesage, S., Brice, A., 2009. Parkinson's disease: from monogenic forms to genetic susceptibility factors. Hum. Mol. Genet. 18, R48-59. https://doi.org/10.1093/hmg/ddp012

Lesage, S., Dürr, A., Tazir, M., Lohmann, E., Leutenegger, A.-L., Janin, S., Pollak, P., Brice, A., 2006. LRRK2 G2019S as a Cause of Parkinson's Disease in North African Arabs. The New England Journal of Medicine 354, 422–3.

LeWitt, P.A., Hauser, R.A., Grosset, D.G., Stocchi, F., Saint-Hilaire, M.-H., Ellenbogen, A., Leinonen, M., Hampson, N.B., DeFeo-Fraulini, T., Freed, M.I., Kieburtz, K.D., 2016. A randomized trial of inhaled levodopa (CVT-301) for motor fluctuations in Parkinson's disease. Mov Disord. n/a-n/a. https://doi.org/10.1002/mds.26611

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Kierczak, M., Shen, X., Ahsan, M., Carlborg, Ö., Marklund, S., 2013. PASE: a novel method for functional prediction of amino acid substitutions based on physicochemical properties. Front Genet 4. https://doi.org/10.3389/fgene.2013.00021

Li, X., Patel, J.C., Wang, J., Avshalumov, M.V., Nicholson, C., Buxbaum, J.D., Elder, G.A., Rice, M.E., Yue, Z., 2010. Enhanced Striatal Dopamine Transmission and Motor Performance with LRRK2 Overexpression in Mice Is Eliminated by Familial Parkinson's Disease Mutation G2019S. J. Neurosci. 30, 1788–1797. https://doi.org/10.1523/JNEUROSCI.5604-09.2010

Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., Benabid, A.-L., 1998. Electrical Stimulation of the Subthalamic Nucleus in Advanced Parkinson's Disease. New England Journal of Medicine 339, 1105–1111. https://doi.org/10.1056/NEJM199810153391603

Liou, H.H., Tsai, M.C., Chen, C.J., Jeng, J.S., Chang, Y.C., Chen, S.Y., Chen, R.C., 1997. Environmental risk factors and Parkinson's disease: a case-control study in Taiwan. Neurology 48, 1583–1588.

Liu, X., Han, S., Wang, Z., Gelernter, J., Yang, B.-Z., 2013. Variant Callers for Next-Generation Sequencing Data: A Comparison Study. PLoS One 8. https://doi.org/10.1371/journal.pone.0075619

Lombard A, Gelfand M, 1978. Parkinson's disease in the African. The Central African journal of medicine 24, 5–8.

Mahne, A.C., Carr, J.A., Bardien, S., Schutte, C.-M., 2016. Clinical findings and genetic screening for copy number variation mutations in a cohort of South African patients with Parkinson's disease. South African Medical Journal 106, 623+.

Mancia, F., Love, J., 2010. High-throughput expression and purification of membrane proteins. J. Struct. Biol. 172, 85–93. https://doi.org/10.1016/j.jsb.2010.03.021

Mandemakers, W., Quadri, M., Stamelou, M., Bonifati, V., 2017. TMEM230: How does it fit in the etiology and pathogenesis of Parkinson's disease? Movement Disorders 32, 1159–1162. https://doi.org/10.1002/mds.27061

Mardis, E.R., 2008. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9, 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Mather, C.A., Mooney, S.D., Salipante, S.J., Scroggins, S., Wu, D., Pritchard, C.C., Shirts, B.H., 2016. CADD score has limited clinical validity for the identification of pathogenic

variants in non-coding regions in a hereditary cancer panel. Genet Med 18, 1269–1275. https://doi.org/10.1038/gim.2016.44

Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. 74, 560–564.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Meireles, J., Massano, J., 2012. Cognitive Impairment and Dementia in Parkinson's Disease: Clinical Features, Diagnosis, and Management. Front Neurol 3. https://doi.org/10.3389/fneur.2012.00088

Meng, H., Yamashita, C., Shiba-Fukushima, K., Inoshita, T., Funayama, M., Sato, S., Hatta, T., Natsume, T., Umitsu, M., Takagi, J., Imai, Y., Hattori, N., 2017a. Loss of Parkinson's disease-associated protein CHCHD2 affects mitochondrial crista structure and destabilizes cytochrome c. Nature Communications 8, 15500. https://doi.org/10.1038/ncomms15500

Meng, H., Yamashita, C., Shiba-Fukushima, K., Inoshita, T., Funayama, M., Sato, S., Hatta, T., Natsume, T., Umitsu, M., Takagi, J., Imai, Y., Hattori, N., 2017b. Loss of Parkinson's disease-associated protein CHCHD2 affects mitochondrial crista structure and destabilizes cytochrome c. Nature Communications 8, 15500. https://doi.org/10.1038/ncomms15500

Metzker, M.L., 2010a. Sequencing technologies — the next generation. Nature Reviews Genetics 11, 31–46. https://doi.org/10.1038/nrg2626

Metzker, M.L., 2010b. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46. https://doi.org/10.1038/nrg2626

Meynert, A.M., Ansari, M., FitzPatrick, D.R., Taylor, M.S., 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 15, 247. https://doi.org/10.1186/1471-2105-15-247

Mi, H., Thomas, P., 2009. PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools, in: Protein Networks and Pathway Analysis, Methods in Molecular Biology. Humana Press, pp. 123–140. https://doi.org/10.1007/978-1-60761-175-2_7

Missler, M., Südhof, T.C., 1998. Neurexins: three genes and 1001 products. Trends Genet. 14, 20–26. https://doi.org/10.1016/S0168-9525(97)01324-3

Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R.E., Gottmann, K., Südhof, T.C., 2003. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature 423, 939–948. https://doi.org/10.1038/nature01755

Møller, R.S., Weber, Y.G., Klitten, L.L., Trucks, H., Muhle, H., Kunz, W.S., Mefford, H.C., Franke, A., Kautza, M., Wolf, P., Dennig, D., Schreiber, S., Rückert, I.-M., Wichmann, H.-E., Ernst, J.P., Schurmann, C., Grabe, H.J., Tommerup, N., Stephani, U., Lerche, H., Hjalgrim, H., Helbig, I., Sander, T., EPICURE Consortium, 2013. Exon-disrupting deletions of NRXN1 in idiopathic generalized epilepsy. Epilepsia 54, 256–264. https://doi.org/10.1111/epi.12078

Moore, D.J., West, A.B., Dawson, V.L., Dawson, T.M., 2005. Molecular pathophysiology of Parkinson's disease. Annu. Rev. Neurosci. 28, 57–87. https://doi.org/10.1146/annurev.neuro.28.061604.135718

Moreland, J.L., Gramada, A., Buzko, O.V., Zhang, Q., Bourne, P.E., 2005. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. BMC Bioinformatics 6, 21. https://doi.org/10.1186/1471-2105-6-21

Mukherjee, K., Sharma, M., Urlaub, H., Bourenkov, G.P., Jahn, R., Südhof, T.C., Wahl, M.C., 2008. CASK Functions as a Mg2+-Independent Neurexin Kinase. Cell 133, 328–339. https://doi.org/10.1016/j.cell.2008.02.036

Müller, W.E., Pedigo, N.W., 1994. Brain aging: a risk factor of neurodegenerative disorders and a target for therapeutic intervention. Life Sci. 55, 1975–1976.

Munoz, D.G., Fujioka, S., 2018. Caffeine and Parkinson disease: A possible diagnostic and pathogenic breakthrough. Neurology 90, 205–206. https://doi.org/10.1212/WNL.0000000000004898

Myhre, R., Steinkjer, S., Stormyr, A., Nilsen, G.L., Zayyad, H.A., Horany, K., Nusier, M.K., Klungland, H., 2008. Significance of the parkin and PINK1 gene in Jordanian families with incidences of young-onset and juvenile parkinsonism. BMC Neurol 8, 47. https://doi.org/10.1186/1471-2377-8-47

Ng M. C. Y., Lam V. K. L., Tam C. H. T., Chan A. W. H., So W.-Y., Ma R. C. W., Zee B. C. Y., Waye M. M. Y., Mak W. W., Hu C., Wang C. R., Tong P. C. Y., Jia W. P., Chan J. C. N., null null, 2010. Association of the POU class 2 homeobox 1 gene (POU2F1) with

susceptibility to Type 2 diabetes in Chinese populations. Diabetic Medicine 27, 1443–1449. https://doi.org/10.1111/j.1464-5491.2010.03124.x

Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J., Bamshad, M.J., 2010. Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. 42, 30–35. https://doi.org/10.1038/ng.499

Nichols, W.C., Pankratz, N., Hernandez, D., Paisán-Ruíz, C., Jain, S., Halter, C.A., Michaels, V.E., Reed, T., Rudolph, A., Shults, C.W., Singleton, A., Foroud, T., Parkinson Study Group-PROGENI investigators, 2005. Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. Lancet 365, 410–412. https://doi.org/10.1016/S0140-6736(05)17828-3

Nisheeth Bandaru, Chen Guo, Yanxi Chen, Tyrone Lee, Peijue Zhang, 2015. In silico prediction of clinical pathogenicty by CADD scoring of exome variants found in genome of a Male belonging to the Chinese Dai Minority. Biology Computes.

Nuytemans, K., Theuns, J., Cruts, M., Van Broeckhoven, C., 2010. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. Hum. Mutat. 31, 763–780. https://doi.org/10.1002/humu.21277

Obeso, J.A., Olanow, C.W., Nutt, J.G., 2000. Levodopa motor complications in Parkinson's disease. Trends in Neurosciences 23, S2–S7. https://doi.org/10.1016/S1471-1931(00)00031-8

Okubadejo, N., Britton, A., Crews, C., Akinyemi, R., Hardy, J., Singleton, A., Bras, J., 2008. Analysis of Nigerians with Apparently Sporadic Parkinson Disease for Mutations in LRRK2 , PRKN and ATXN3. PLOS ONE 3, e3421. https://doi.org/10.1371/journal.pone.0003421

Okubadejo, N.U., 2008a. An analysis of genetic studies of Parkinson's disease in Africa. Parkinsonism & Related Disorders 14, 177–182. https://doi.org/10.1016/j.parkreldis.2007.08.006

Okubadejo, N.U., 2008b. An analysis of genetic studies of Parkinson's disease in Africa. Parkinsonism & Related Disorders 14, 177–182. https://doi.org/10.1016/j.parkreldis.2007.08.006

Osuntokun, B.., Adeuja, A.O.., Schoenberg, B.., Bademosi, O., Nottidge, V.., Olumide, A.., Lge, O., Yaria, F., Bolis, C.., 1987. Neurological disorders in Nigerian Africans: a community-based study. Acta Neurologica Scandinavica 75, 13–21. https://doi.org/10.1111/j.1600-0404.1987.tb07883.x

Ovallath, S., Deepa, P., 2013. The history of parkinsonism: Descriptions in ancient Indian medical literature. Mov Disord. 28, 566–568. https://doi.org/10.1002/mds.25420

Paisán-Ruíz, C., Jain, S., Evans, E.W., Gilks, W.P., Simón, J., van der Brug, M., López de Munain, A., Aparicio, S., Gil, A.M., Khan, N., Johnson, J., Martinez, J.R., Nicholl, D., Carrera, I.M., Pena, A.S., de Silva, R., Lees, A., Martí-Massó, J.F., Pérez-Tur, J., Wood, N.W., Singleton, A.B., 2004. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44, 595–600. https://doi.org/10.1016/j.neuron.2004.10.023

Parkinson, J., 1817. An Essay on the Shaking Palsy. Sheerwood, Neely and Jones, Paternoster Row , London.

Patel, Z.H., Kottyan, L.C., Lazaro, S., Williams, M.S., Ledbetter, D.H., Tromp, H.B.,Gerard, Rupert, A., Kohram, M., Wagner, M., Husami, A., Qian, Y., Valencia, C.A., Zhang, K., Hostetter, M.K., Harley, J.B., Kaufman, K.M., 2014. The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors. Front Genet 5. https://doi.org/10.3389/fgene.2014.00016

Perier, C., Bové, J., Vila, M., 2011. Mitochondria and Programmed Cell Death in Parkinson's Disease: Apoptosis and Beyond. Antioxidants & Redox Signaling 16, 883–895. https://doi.org/10.1089/ars.2011.4074

Picconi, B., Piccoli, G., Calabresi, P., 2012. Synaptic dysfunction in Parkinson's disease. Adv. Exp. Med. Biol. 970, 553–572. https://doi.org/10.1007/978-3-7091-0932-8_24

Pillai, S., Gopalan, V., Lam, A.K.-Y., 2017. Review of sequencing platforms and their applications in phaeochromocytoma and paragangliomas. Critical Reviews in Oncology / Hematology 116, 58–67. https://doi.org/10.1016/j.critrevonc.2017.05.005

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., Siepel, A., 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20, 110–121. https://doi.org/10.1101/gr.097857.109

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E.S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W.G., Lazzarini, A.M., Duvoisin, R.C., Di Iorio, G., Golbe, L.I., Nussbaum, R.L., 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276, 2045–2047.

Pontén F., Schwenk J. M., Asplund A., Edqvist P.-H. D., 2011. The Human Protein Atlas as a proteomic resource for biomarker discovery. Journal of Internal Medicine 270, 428–446. https://doi.org/10.1111/j.1365-2796.2011.02427.x

Postuma, R.B., Lang, A.E., Munhoz, R.P., Charland, K., Pelletier, A., Moscovich, M., Filla, L., Zanatta, D., Rios Romenets, S., Altman, R., Chuang, R., Shah, B., 2012. Caffeine for treatment of Parkinson disease: a randomized controlled trial. Neurology 79, 651–658. https://doi.org/10.1212/WNL.0b013e318263570d

Pringsheim, T., Jette, N., Frolkis, A., Steeves, T.D.L., 2014a. The prevalence of Parkinson's disease: A systematic review and meta-analysis. Mov Disord. 29, 1583–1590. https://doi.org/10.1002/mds.25945

Pringsheim, T., Jette, N., Frolkis, A., Steeves, T.D.L., 2014b. The prevalence of Parkinson's disease: A systematic review and meta-analysis. Movement Disorders 29, 1583–1590. https://doi.org/10.1002/mds.25945

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., Murphy, M.R., O'Leary, N.A., Pujar, S., Rajput, B., Rangwala, S.H., Riddick, L.D., Shkeda, A., Sun, H., Tamez, P., Tully, R.E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D.R., Murphy, T.D., Ostell, J.M., 2014. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 42, D756-763. https://doi.org/10.1093/nar/gkt1114

Quadri, M., Fang, M., Picillo, M., Olgiati, S., Breedveld, G.J., Graafland, J., Wu, B., Xu, F., Erro, R., Amboni, M., Pappatà, S., Quarantelli, M., Annesi, G., Quattrone, A., Chien, H.F., Barbosa, E.R., International Parkinsonism Genetics Network, Oostra, B.A., Barone, P., Wang, J., Bonifati, V., 2013. Mutation in the SYNJ1 gene associated with autosomal recessive, early-onset Parkinsonism. Hum. Mutat. 34, 1208–1215. https://doi.org/10.1002/humu.22373

Quadri, M., Mandemakers, W., Grochowska, M.M., Masius, R., Geut, H., Fabrizio, E., Breedveld, G.J., Kuipers, D., Minneboo, M., Vergouw, L.J.M., Carreras Mascaro, A., Yonova-Doing, E., Simons, E., Zhao, T., Di Fonzo, A.B., Chang, H.-C., Parchi, P., Melis, M., Correia Guedes, L., Criscuolo, C., Thomas, A., Brouwer, R.W.W., Heijsman, D., Ingrassia, A.M.T., Calandra Buonaura, G., Rood, J.P., Capellari, S., Rozemuller, A.J., Sarchioto, M., Fen Chien, H., Vanacore, N., Olgiati, S., Wu-Chou, Y.-H., Yeh, T.-H., Boon, A.J.W., Hoogers, S.E., Ghazvini, M., IJpma, A.S., van IJcken, W.F.J., Onofrj, M., Barone, P., Nicholl, D.J., Puschmann, A., De Mari, M., Kievit, A.J., Barbosa, E., De Michele, G., Majoor-Krakauer, D., van Swieten, J.C., de Jong, F.J., Ferreira, J.J., Cossu, G., Lu, C.-S., Meco, G., Cortelli, P., van de Berg, W.D.J., Bonifati, V., 2018a. LRP10 genetic variants in familial Parkinson's disease and dementia with Lewy bodies: a genome-wide linkage and sequencing study. The Lancet Neurology 17, 597–608. https://doi.org/10.1016/S1474-4422(18)30179-0

Quadri, M., Mandemakers, W., Grochowska, M.M., Masius, R., Geut, H., Fabrizio, E., Breedveld, G.J., Kuipers, D., Minneboo, M., Vergouw, L.J.M., Carreras Mascaro, A., Yonova-Doing, E., Simons, E., Zhao, T., Di Fonzo, A.B., Chang, H.-C., Parchi, P., Melis, M., Correia Guedes, L., Criscuolo, C., Thomas, A., Brouwer, R.W.W., Heijsman, D., Ingrassia, A.M.T., Calandra Buonaura, G., Rood, J.P., Capellari, S., Rozemuller, A.J., Sarchioto, M., Fen Chien, H., Vanacore, N., Olgiati, S., Wu-Chou, Y.-H., Yeh, T.-H., Boon, A.J.W., Hoogers, S.E., Ghazvini, M., IJpma, A.S., van IJcken, W.F.J., Onofrj, M., Barone, P., Nicholl, D.J., Puschmann, A., De Mari, M., Kievit, A.J., Barbosa, E., De Michele, G., Majoor-Krakauer, D., van Swieten, J.C., de Jong, F.J., Ferreira, J.J., Cossu, G., Lu, C.-S., Meco, G., Cortelli, P., van de Berg, W.D.J., Bonifati, V., International Parkinsonism Genetics Network, 2018b. LRP10 genetic variants in familial Parkinson's disease and dementia with Lewy bodies: a genome-wide linkage and sequencing study. Lancet Neurol 17, 597–608. https://doi.org/10.1016/S1474-4422(18)30179-0

Quik, M., Kulak, J.M., 2002. Nicotine and Nicotinic Receptors; Relevance to Parkinson's Disease. NeuroToxicology 23, 581–594. https://doi.org/10.1016/S0161-813X(02)00036-0

Quinlan, K.G.R., Seto, J.T., Turner, N., Vandebrouck, A., Floetenmeyer, M., Macarthur, D.G., Raftery, J.M., Lek, M., Yang, N., Parton, R.G., Cooney, G.J., North, K.N., 2010. α-Actinin-3 deficiency results in reduced glycogen phosphorylase activity and altered calcium handling in skeletal muscle. Hum Mol Genet 19, 1335–1346. https://doi.org/10.1093/hmg/ddq010

Rabbani, B., Tekin, M., Mahdieh, N., 2014. The promise of whole-exome sequencing in medical genetics. J. Hum. Genet. 59, 5–15. https://doi.org/10.1038/jhg.2013.114

Ragland, M., Hutter, C., Zabetian, C., Edwards, K., 2009. Association Between the Ubiquitin Carboxyl-Terminal Esterase L1 Gene (UCHL1) S18Y Variant and Parkinson's Disease: A HuGE Review and Meta-Analysis. Am J Epidemiol 170, 1344–1357. https://doi.org/10.1093/aje/kwp288

Ramsay, M., Tiemessen, C.T., Choudhury, A., Soodyall, H., 2011. Africa: the next frontier for human disease gene discovery? Hum. Mol. Genet. 20, R214–R220. https://doi.org/10.1093/hmg/ddr401

Reeve, A., Simcox, E., Turnbull, D., 2014. Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? Ageing Research Reviews 14, 19–30. https://doi.org/10.1016/j.arr.2014.01.004

Reissner, C., Klose, M., Fairless, R., Missler, M., 2008. Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. PNAS 105, 15124–15129. https://doi.org/10.1073/pnas.0801639105

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17, 405–424. https://doi.org/10.1038/gim.2015.30

Ritchie, G.R.S., Dunham, I., Zeggini, E., Flicek, P., 2014. Functional annotation of noncoding sequence variants. Nat. Methods 11, 294–296. https://doi.org/10.1038/nmeth.2832

Roach, E.S., 2007. Both Postsynaptic and Presynaptic Dysfunction Contribute to Parkinson Disease: Any Mechanism Will Not Do. Arch Neurol 64, 143–143. https://doi.org/10.1001/archneur.64.1.143

Robert, C., Fuentes-Utrilla, P., Troup, K., Loecherbach, J., Turner, F., Talbot, R., Archibald, A.L., Mileham, A., Deeb, N., Hume, D.A., Watson, M., 2014. Design and development of exome capture sequencing for the domestic pig (Sus scrofa). BMC Genomics 15, 550. https://doi.org/10.1186/1471-2164-15-550

Rogaeva, E., Johnson, J., Lang, A.E., Gulick, C., Gwinn-Hardy, K., Kawarai, T., Sato, C., Morgan, A., Werner, J., Nussbaum, R., Petit, A., Okun, M.S., McInerney, A., Mandel, R.,

Groen, J.L., Fernandez, H.H., Postuma, R., Foote, K.D., Salehi-Rad, S., Liang, Y., Reimsnider, S., Tandon, A., Hardy, J., St George-Hyslop, P., Singleton, A.B., 2004. Analysis of the PINK1 gene in a large cohort of cases with Parkinson disease. Arch. Neurol. 61, 1898–1904. https://doi.org/10.1001/archneur.61.12.1898

Ross, G.W., Abbott, R.D., Petrovitch, H., Morens, D.M., Grandinetti, A., Tung, K.H., Tanner, C.M., Masaki, K.H., Blanchette, P.L., Curb, J.D., Popper, J.S., White, L.R., 2000. Association of coffee and caffeine intake with the risk of Parkinson disease. JAMA 283, 2674–2679.

Rudenko, G., Hohenester, E., Muller, Y.A., 2001. LG/LNS domains: multiple functions – one business end? Trends in Biochemical Sciences 26, 363–368. https://doi.org/10.1016/S0968-0004(01)01832-1

Rudenko, G., Nguyen, T., Chelliah, Y., Sudhof, T.C., Deisenhofer, J., 1999. The structure of the ligand-binding domain of neurexin Ib: regulation of LNS domain function by alternative splicing. Cell 99, 93–102.

Samaranch, L., Lorenzo-Betancor, O., Arbelo, J.M., Ferrer, I., Lorenzo, E., Irigoyen, J., Pastor, M.A., Marrero, C., Isla, C., Herrera-Henriquez, J., Pastor, P., 2010. PINK1-linked parkinsonism is associated with Lewy body pathology. Brain 133, 1128–1142. https://doi.org/10.1093/brain/awq051

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Saunders-Pullman, R., 2003. Estrogens and Parkinson disease: Neuroprotective, symptomatic, neither, or both? Endocrine 21, 81–87. https://doi.org/10.1385/ENDO:21:1:81

Saux, O.L., Beck, K., Sachsinger, C., Treiber, C., Göring, H.H., Curry, K., Johnson, E.W., Bercovitch, L., Marais, A.-S., Terry, S.F., Viljoen, D.L., Boyd, C.D., 2002. Evidence for a founder effect for pseudoxanthoma elasticum in the Afrikaner population of South Africa. Hum Genet 111, 331–338. https://doi.org/10.1007/s00439-002-0808-1

Schaaf, C.P., Boone, P.M., Sampath, S., Williams, C., Bader, P.I., Mueller, J.M., Shchelochkov, O.A., Brown, C.W., Crawford, H.P., Phalen, J.A., Tartaglia, N.R., Evans, P., Campbell, W.M., Chun-Hui Tsai, A., Parsley, L., Grayson, S.W., Scheuerle, A., Luzzi, C.D., Thomas, S.K., Eng, P.A., Kang, S.-H.L., Patel, A., Stankiewicz, P., Cheung, S.W., 2012. Phenotypic spectrum and genotype–phenotype correlations of NRXN1 exon deletions. Eur J Hum Genet 20, 1240–1247. https://doi.org/10.1038/ejhg.2012.95

Schapira, A.H., Jenner, P., 2011. Etiology and pathogenesis of Parkinson's disease. Mov. Disord. 26, 1049–1055. https://doi.org/10.1002/mds.23732

Schoenberg, B.S., Osuntokun, B.O., Adeuja, A.O.G., Bademosi, O., Nottidge, V., Anderson, D.W., Haerer, A.F., 1988. Comparison of the prevalence of Parkinson's disease in black populations in the rural United States and in rural Nigeria Door-to-door community studies. Neurology 38, 645–645. https://doi.org/10.1212/WNL.38.4.645

Schreiner, D., Nguyen, T.-M., Russo, G., Heber, S., Patrignani, A., Ahrné, E., Scheiffele, P., 2014. Targeted Combinatorial Alternative Splicing Generates Brain Region-Specific Repertoires of Neurexins. Neuron 84, 386–398. https://doi.org/10.1016/j.neuron.2014.09.011

Schreiner, D., Simicevic, J., Ahrné, E., Schmidt, A., Scheiffele, P., 2015. Quantitative isoform-profiling of highly diversified recognition molecules. eLife 4. https://doi.org/10.7554/eLife.07794

Schrödinger L, 2015. The PyMOL Molecular Graphics System, Version 2.0 [WWW Document]. URL http://www.pymol.org/ (accessed 10.24.18).

Schulte, E.C., Mollenhauer, B., Zimprich, A., Bereznai, B., Lichtner, P., Haubenberger, D., Pirker, W., Brücke, T., Molnar, M.J., Peters, A., Gieger, C., Trenkwalder, C., Winkelmann, J., 2012. Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease. Neurogenetics 13, 281–285. https://doi.org/10.1007/s10048-012-0334-9

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. Nat. Methods 5, 16–18. https://doi.org/10.1038/nmeth1156

Schwarz, J.M., Cooper, D.N., Schuelke, M., Seelow, D., 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Meth 11, 361–362. https://doi.org/10.1038/nmeth.2890

Schwarze, K., Buchanan, J., Taylor, J.C., Wordsworth, S., 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genetics in Medicine. https://doi.org/10.1038/gim.2017.247

Seidler, A., Hellenbrand, W., Robra, B.P., Vieregge, P., Nischan, P., Joerg, J., Oertel, W.H., Ulm, G., Schneider, E., 1996. Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. Neurology 46, 1275–1284.

Semchuk, K.M., Love, E.J., Lee, R.G., 1993. Parkinson's disease: a test of the multifactorial etiologic hypothesis. Neurology 43, 1173–1180.

Semchuk, K.M., Love, E.J., Lee, R.G., 1992. Parkinson's disease and exposure to agricultural work and pesticide chemicals. Neurology 42, 1328–1335.

Semchuk, K.M., Love, E.J., Lee, R.G., 1991. Parkinson's disease and exposure to rural environmental factors: a population based case-control study. Canadian journal of neurological sciences 18, 279–286.

Sherer, T.B., Richardson, J.R., Testa, C.M., Seo, B.B., Panov, A.V., Yagi, T., Matsuno-Yagi, A., Miller, G.W., Greenamyre, J.T., 2007. Mechanism of toxicity of pesticides acting at complex I: relevance to environmental etiologies of Parkinson's disease. J. Neurochem. 100, 1469–1479. https://doi.org/10.1111/j.1471-4159.2006.04333.x

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29, 308–311.

Shihab, H.A., Rogers, M.F., Gough, J., Mort, M., Cooper, D.N., Day, I.N.M., Gaunt, T.R., Campbell, C., 2015. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 31, 1536–1543. https://doi.org/10.1093/bioinformatics/btv009

Siddiqui, T.J., Pancaroglu, R., Kang, Y., Rooyakkers, A., Craig, A.M., 2010. LRRTMs and Neuroligins Bind Neurexins with a Differential Code to Cooperate in Glutamate Synapse Development. J. Neurosci. 30, 7495–7506. https://doi.org/10.1523/JNEUROSCI.0470-10.2010

Sidransky, E., Lopez, G., 2012. The link between the GBA gene and parkinsonism. The Lancet Neurology 11, 986–998. https://doi.org/10.1016/S1474-4422(12)70190-4

Siepel, A., 2005. PhastCons HOWTO [WWW Document]. URL http://compgen.cshl.edu/phast/phastCons-HOWTO.html (accessed 7.5.18).

Siva, N., 2008. 1000 Genomes project [WWW Document]. Nature Biotechnology. https://doi.org/10.1038/nbt0308-256b

Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M., Goedert, M., 1998. α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. Proc Natl Acad Sci U S A 95, 6469–6473.

Stefani, A., Lozano, A.M., Peppe, A., Stanzione, P., Galati, S., Tropepi, D., Pierantozzi, M., Brusa, L., Scarnati, E., Mazzone, P., 2007. Bilateral deep brain stimulation of the

pedunculopontine and subthalamic nuclei in severe Parkinson's disease. Brain 130, 1596–1607. https://doi.org/10.1093/brain/awl346

Steinberg, F., Gallon, M., Winfield, M., Thomas, E., Bell, A.J., Heesom, K.J., Tavaré, J.M., Cullen, P.J., 2013. A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose and metal ion transport. Nat Cell Biol 15, 461–471. https://doi.org/10.1038/ncb2721

Sudhaman, S., Prasad, K., Behari, M., Muthane, U.B., Juyal, R.C., Thelma, B.K., 2016. Discovery of a frameshift mutation in podocalyxin-like (PODXL) gene, coding for a neural adhesion molecule, as causal for autosomal-recessive juvenile Parkinsonism. J. Med. Genet. 53, 450–456. https://doi.org/10.1136/jmedgenet-2015-103459

Sun, Q., Wang, Y., Zhang, Y., Liu, F., Cheng, X., Hou, N., Zhao, X., Yang, X., 2007. Expression profiling reveals dysregulation of cellular cytoskeletal genes in HBx–induced hepatocarcinogenesis. Cancer Biology & Therapy 6, 668–674. https://doi.org/10.4161/cbt.6.5.3955

Sunkin, S.M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T.L., Thompson, C.L., Hawrylycz, M., Dang, C., 2013. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res 41, D996–D1008. https://doi.org/10.1093/nar/gks1042

Tabuchi, K., Südhof, T.C., 2002. Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. Genomics 79, 849–859. https://doi.org/10.1006/geno.2002.6780

Tan, E.-K., Yew, K., Chua, E., Puvan, K., Shen, H., Lee, E., Puong, K.-Y., Zhao, Y., Pavanni, R., Wong, M.-C., Jamora, D., de Silva, D., Moe, K.-T., Woon, F.-P., Yuen, Y., Tan, L., 2006. PINK1 mutations in sporadic early-onset Parkinson's disease. Mov. Disord. 21, 789–793. https://doi.org/10.1002/mds.20810

Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, Jiaxiong, Zhang, L., Sun, X., Zhi, E., Wang, H., Li, H., Jin, L., Luo, Y., Wang, Jian, Yang, S., Zhang, F., 2017. Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. The American Journal of Human Genetics 100, 854–864. https://doi.org/10.1016/j.ajhg.2017.04.012

Tattini, L., D'Aurizio, R., Magi, A., 2015. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Front Bioeng Biotechnol 3. https://doi.org/10.3389/fbioe.2015.00092

Thacker, E.L., O'Reilly, E.J., Weisskopf, M.G., Chen, H., Schwarzschild, M.A., McCullough, M.L., Calle, E.E., Thun, M.J., Ascherio, A., 2007. Temporal relationship between cigarette smoking and risk of Parkinson disease. Neurology 68, 764–768. https://doi.org/10.1212/01.wnl.0000256374.50227.4b

Thal, D.R., Del Tredici, K., Braak, H., 2004. Neurodegeneration in normal brain aging and disease. Sci Aging Knowledge Environ 2004, pe26. https://doi.org/10.1126/sageke.2004.23.pe26

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393

Tipping, A.J., Pearson, T., Morgan, N.V., Gibson, R.A., Kuyt, L.P., Havenga, C., Gluckman, E., Joenje, H., de Ravel, T., Jansen, S., Mathew, C.G., 2001. Molecular and genealogical evidence for a founder effect in Fanconi anemia families of the Afrikaner population of South Africa. Proceedings of the National Academy of Sciences 98, 5734–5739. https://doi.org/10.1073/pnas.091402398

Tolosa, E., Wenning, G., Poewe, W., 2006. The diagnosis of Parkinson's disease. The Lancet Neurology 5, 75–86. https://doi.org/10.1016/S1474-4422(05)70285-4

Topper, L.M., Campbell, M.S., Tugendreich, S., Daum, J.R., Burke, D.J., Hieter, P., Gorbsky, G.J., 2002. The dephosphorylated form of the anaphase-promoting complex protein Cdc27/Apc3 concentrates on kinetochores and chromosome arms in mitosis. Cell Cycle 1, 282–292.

Trenkwalder, C., Arnulf, I., Postuma, R., 2017. Parkinsonism, in: Principles and Practice of Sleep Medicine. Elsevier, pp. 892-902.e5. https://doi.org/10.1016/B978-0-323-24288-2.00092-1

Trinh, J., Guella, I., McKenzie, M., Gustavsson, E.K., Szu-Tu, C., Petersen, M.S., Rajput, A., Rajput, A.H., McKeown, M., Jeon, B.S., Aasly, J.O., Bardien, S., Farrer, M.J., 2015. Novel LRRK2 mutations in Parkinsonism. Parkinsonism & Related Disorders 21, 1119–1121. https://doi.org/10.1016/j.parkreldis.2015.07.011

Uemura, T., Lee, S.-J., Yasumura, M., Takeuchi, T., Yoshida, T., Ra, M., Taguchi, R., Sakimura, K., Mishina, M., 2010. Trans-Synaptic Interaction of GluRδ2 and Neurexin through Cbln1 Mediates Synapse Formation in the Cerebellum. Cell 141, 1068–1079. https://doi.org/10.1016/j.cell.2010.04.035

Ullrich, B., Ushkaryov, Y.A., Südhof, T.C., 1995. Cartography of neurexins: More than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron 14, 497–507. https://doi.org/10.1016/0896-6273(95)90306-2

Ushkaryov, Y.A., Petrenko, A.G., Geppert, M., Sudhof, T.C., 1992. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science 257, 50–56. https://doi.org/10.1126/science.1621094

Ushkaryov, Y.A., Südhof, T.C., 1993. Neurexin III alpha: extensive alternative splicing generates membrane-bound and soluble forms. PNAS 90, 6410–6414. https://doi.org/10.1073/pnas.90.14.6410

Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M.K., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., Albanese, A., Nussbaum, R., González-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W.P., Latchman, D.S., Harvey, R.J., Dallapiccola, B., Auburger, G., Wood, N.W., 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304, 1158–1160. https://doi.org/10.1126/science.1096284

van der Merwe, C., Carr, J., Glanzmann, B., Bardien, S., 2016. Exonic rearrangements in the known Parkinson's disease-causing genes are a rare cause of the disease in South African patients. Neuroscience Letters 619, 168–171. https://doi.org/10.1016/j.neulet.2016.03.028

Velkoff, V.A., Kowal, P.R., Census Bureau, U.S., 2007. Population Aging in Sub-Saharan Africa: Demographic Dimensions 2006. U.S. Government Printing Office,Washington, DC. https://doi.org/P95/07-1

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y.,

Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304–1351. https://doi.org/10.1126/science.1058040

Verstraeten, A., Theuns, J., Van Broeckhoven, C., 2015. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. Trends in Genetics 31, 140–149. https://doi.org/10.1016/j.tig.2015.01.004

Vilariño-Güell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., Behrouz, B., Melrose, H.L., Hentati, E., Puschmann, A., Evans, D.M., Conibear, E., Wasserman, W.W., Aasly, J.O.,

Burkhard, P.R., Djaldetti, R., Ghika, J., Hentati, F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A.H., Solida, A., Wu, R.-M., Uitti, R.J., Wszolek, Z.K., Vingerhoets, F., Farrer, M.J., 2011a. VPS35 Mutations in Parkinson Disease. Am J Hum Genet 89, 162–167. https://doi.org/10.1016/j.ajhg.2011.06.001

Vilariño-Güell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., Behrouz, B., Melrose, H.L., Hentati, E., Puschmann, A., Evans, D.M., Conibear, E., Wasserman, W.W., Aasly, J.O., Burkhard, P.R., Djaldetti, R., Ghika, J., Hentati, F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A.H., Solida, A., Wu, R.-M., Uitti, R.J., Wszolek, Z.K., Vingerhoets, F., Farrer, M.J., 2011b. VPS35 Mutations in Parkinson Disease. Am J Hum Genet 89, 162–167. https://doi.org/10.1016/j.ajhg.2011.06.001

Wang, L., Aasly, J.O., Annesi, G., Bardien, S., Bozi, M., Brice, A., Carr, J., Chung, S.J., Clarke, C., Crosiers, D., Deutschländer, A., Eckstein, G., Farrer, M.J., Goldwurm, S., Garraux, G., Hadjigeorgiou, G.M., Hicks, A.A., Hattori, N., Klein, C., Jeon, B., Kim, Y.J., Lesage, S., Lin, J.-J., Lynch, T., Lichtner, P., Lang, A.E., Mok, V., Jasinska-Myga, B., Mellick, G.D., Morrison, K.E., Opala, G., Pihlstrøm, L., Pramstaller, P.P., Park, S.S., Quattrone, A., Rogaeva, E., Ross, O.A., Stefanis, L., Stockton, J.D., Silburn, P.A., Theuns, J., Tan, E.K., Tomiyama, H., Toft, M., Van Broeckhoven, C., Uitti, R.J., Wirdefeldt, K., Wszolek, Z., Xiromerisiou, G., Yueh, K.-C., Zhao, Y., Gasser, T., Maraganore, D.M., Krüger, R., Sharma, M., 2015. Large-scale assessment of polyglutamine repeat expansions in Parkinson disease. Neurology 85, 1283–1292. https://doi.org/10.1212/WNL.0000000000002016

Watanabe, Y., Himeda, T., Araki, T., 2005. Mechanisms of MPTP toxicity and their implications for therapy of Parkinson's disease. Med. Sci. Monit. 11, RA17-23.

Weaver, F.M., 2009. Bilateral Deep Brain Stimulation vs Best Medical Therapy for Patients With Advanced Parkinson DiseaseA Randomized Controlled Trial. JAMA 301, 63. https://doi.org/10.1001/jama.2008.929

Weinrich, S.L., Pruzan, R., Ma, L., Ouellette, M., Tesmer, V.M., Holt, S.E., Bodnar, A.G., Lichtsteiner, S., Kim, N.W., Trager, J.B., Taylor, R.D., Carlos, R., Andrews, W.H., Wright, W.E., Shay, J.W., Harley, C.B., Morin, G.B., 1997. Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. Nat. Genet. 17, 498–502. https://doi.org/10.1038/ng1297-498

Williams, K.L., Topp, S., Yang, S., Smith, B., Fifita, J.A., Warraich, S.T., Zhang, K.Y., Farrawell, N., Vance, C., Hu, X., Chesi, A., Leblond, C.S., Lee, A., Rayner, S.L., Sundaramoorthy, V., Dobson-Stone, C., Molloy, M.P., Blitterswijk, M. van, Dickson, D.W., Petersen, R.C., Graff-Radford, N.R., Boeve, B.F., Murray, M.E., Pottier, C., Don, E., Winnick, C., McCann, E.P., Hogan, A., Daoud, H., Levert, A., Dion, P.A., Mitsui, J., Ishiura, H., Takahashi, Y., Goto, J., Kost, J., Gellera, C., Gkazi, A.S., Miller, J., Stockton, J., Brooks, W.S., Boundy, K., Polak, M., Muñoz-Blanco, J.L., Esteban-Pérez, J., Rábano, A., Hardiman, O., Morrison, K.E., Ticozzi, N., Silani, V., Belleroche, J. de, Glass, J.D., Kwok, J.B.J., Guillemin, G.J., Chung, R.S., Tsuji, S., Jr, R.H.B., García-Redondo, A., Rademakers, R., Landers, J.E., Gitler, A.D., Rouleau, G.A., Cole, N.J., Yerbury, J.J., Atkin, J.D., Shaw, C.E., Nicholson, G.A., Blair, I.P., 2016. *CCNF* mutations in amyotrophic lateral sclerosis and frontotemporal dementia. Nature Communications 7, 11253. https://doi.org/10.1038/ncomms11253

Winkler, A.S., Tütüncü, E., Trendafilova, A., Meindl, M., Kaaya, J., Schmutzhard, E., Kassubek, J., 2010. Parkinsonism in a population of northern Tanzania: a community-based door-to-door study in combination with a prospective hospital-based evaluation. Journal of Neurology 257, 799–805. http://dx.doi.org/10.1007/s00415-009-5420-z

Winklhofer, K.F., Haass, C., 2010. Mitochondrial dysfunction in Parkinson's disease. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, Mitochondrial Dysfunction 1802, 29–44. https://doi.org/10.1016/j.bbadis.2009.08.013

Wong, G.K.-S., Yang, Z., Passey, D.A., Kibukawa, M., Paddock, M., Liu, C.-R., Bolund, L., Yu, J., 2003. A Population Threshold for Functional Polymorphisms. Genome Res 13, 1873–1879. https://doi.org/10.1101/gr.1324303

Yan, W., Tang, B., Zhou, X., Lei, L., Li, K., Sun, Q., Xu, Q., Yan, X., Guo, J., Liu, Z., 2017. TMEM230 mutation analysis in Parkinson's disease in a Chinese population. Neurobiol. Aging 49, 219.e1-219.e3. https://doi.org/10.1016/j.neurobiolaging.2016.10.007

Yang, X., An, R., Xi, J., Zheng, J., Chen, Y., Huang, H., Tian, S., Zhao, Q., Ning, P., Xu, Y., 2017. Sequencing TMEM230 in Chinese patients with sporadic or familial Parkinson's disease. Movement Disorders 32, 800–802. https://doi.org/10.1002/mds.26996

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S.E., Lupski, J.R., Beaudet, A.L., Gibbs, R.A., Eng, C.M., 2013. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian

Disorders. New England Journal of Medicine 369, 1502–1511. https://doi.org/10.1056/NEJMoa1306555

Yangngam, S., Plong-On, O., Sripo, T., Roongpraiwan, R., Hansakunachai, T., Wirojanan, J., Sombuntham, T., Ruangdaraganon, N., Limprasert, P., 2014. Mutation Screening of the Neurexin 1 Gene in Thai Patients with Intellectual Disability and Autism Spectrum Disorder. Genet Test Mol Biomarkers 18, 510–515. https://doi.org/10.1089/gtmb.2014.0003

Yonova-Doing, E., Atadzhanov, M., Quadri, M., Kelly, P., Shawa, N., Musonda, S.T.S., Simons, E.J., Breedveld, G.J., Oostra, B.A., Bonifati, V., 2012. Analysis of LRRK2, SNCA, Parkin, PINK1, and DJ-1 in Zambian patients with Parkinson's disease. Parkinsonism & Related Disorders 18, 567–571. https://doi.org/10.1016/j.parkreldis.2012.02.018

Zarranz, J.J., Alegre, J., Gómez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Atarés, B., Llorens, V., Gomez Tortosa, E., del Ser, T., Muñoz, D.G., de Yebenes, J.G., 2004. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann. Neurol. 55, 164–173. https://doi.org/10.1002/ana.10795

Zarranz, J.J., Fernández-Bedoya, A., Lambarri, I., Gómez-Esteban, J.C., Lezcano, E., Zamacona, J., Madoz, P., 2005. Abnormal sleep architecture is an early feature in the E46K familial synucleinopathy. Mov. Disord. 20, 1310–1315. https://doi.org/10.1002/mds.20581

Zhang, Z.X., Román, G.C., 1993. Worldwide occurrence of Parkinson's disease: an updated review. Neuroepidemiology 12, 195–208. https://doi.org/10.1159/000110318

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., Rossle, S.C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Brücke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T., Strom, T.M., 2011. A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. Am J Hum Genet 89, 168–175. https://doi.org/10.1016/j.ajhg.2011.06.008

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., Stoessl, A.J., Pfeiffer, R.F., Patenge, N., Carbajal, I.C., Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D.W., Meitinger, T., Strom,

T.M., Wszolek, Z.K., Gasser, T., 2004. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44, 601–607. https://doi.org/10.1016/j.neuron.2004.11.005

Zou, Y.-M., Liu, J., Tian, Z.-Y., Lu, D., Zhou, Y.-Y., 2015. Systematic review of the prevalence and incidence of Parkinson's disease in the People's Republic of China. Neuropsychiatr Dis Treat 11, 1467–1472. https://doi.org/10.2147/NDT.S85380

Zrinzo, L., Foltynie, T., Limousin, P., Hariz, M.I., 2011. Reducing hemorrhagic complications in functional neurosurgery: a large case series and systematic literature review. Journal of Neurosurgery 116, 84–94. https://doi.org/10.3171/2011.8.JNS101407

URL List

Institute/Program	Web address
Background	
National Human Genome Research Institute	https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/
Pedigree Chart	
GenoPro Genealogy Software	https://www.genopro.com/
WES Analysis	
Burrows-Wheeler Aligner	bio-bwa.sourceforge.net/
ANNOVAR Software	annovar.openbioinformatics.org/
Genome Analysis Toolkit	https://software.broadinstitute.org/gatk/
SAMtools mpileup	samtools.sourceforge.net/mpileup.shtml
Sanger Sequencing	
Primer3 4.0.0	http://primer3.ut.ee
Primer - Basic Local Alignment Search Tool	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
BioEdit 7.2.0	www.mbio.ncsu.edu/BioEdit/bioedit.html
Ensembl Genome Browser database	http://www.ensembl.org
Bioinformatic In-silico Analysis	
ExAC database	http://exac.broadinstitute.org/
1000 Genomes Project	http://browser.1000genomes.org/index.html

dbSNP	https://www.ncbi.nlm.nih.gov/snp
SIFT	http://sift.jcvi.org/
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
MutationTaster	http://www.mutationtaster.org/)
CADD	http://cadd.gs.washington.edu/
GERP++	http://mendel.stanford.edu/SidowLab/downloads/gerp/
KEGG Pathways Analyser	http://www.genome.jp/kegg/pathway.html
PANTHER Pathway Analyser	http://www.pantherdb.org/pathway/
Allen Brain Atlas	www.brain-map.org/
Human Protein Atlas	https://www.proteinatlas.org/
AnnEx	https://annex.can.ubc.ca/
Protein modelling	
SWISS-MODEL Webserver	https://swissmodel.expasy.org/
BLASTP	https://blast.ncbi.nlm.nih.gov/BLAST.cgi?PAGE=Proteins
PyMOL /	https://pymol.org/
Swiss-PDB Viewer	https://spdbv.vital-it.ch/
Western Blot Analysis	
ImageJ	http://imagej.nih.gov/ij/

Appendices

1. Appendix I

1. Supplementary Tables

Table 1 List of genes not conclusively implicated in PD

Gene (locus)	Typical symptoms	Age of onset	Reference
ATP13A2 **(PARK9)**	Early-onset Dementia Spasticity supranuclear Gaze palsy Fast progression	≤50	Ferreira et al. 2016
ATP1A3	Early-onset Dystonia Poor response to Levodopa Fast progression	NR	McKeon et al. 2007
CHCHD2	Autosomal Dominant Late-onset Hyperreflexia Good response to Levodopa Slow progression	≥60	Funayama et al 2015 Jansen et al., 2015; Koschmidder et al., 2016; Gagliardi et al., 2017 Meng et al., 2017
DCTN1	Late-onset parkinsonism Poor response to Levodopa	≥55	Aji et al. 2013 Araki et al. 2014

Gene (locus)	Typical symptoms	Age of onset	Reference
DNAJC13	Early-onset parkinsonism Lewy bodies present Good response to Levodopa Slow progression	≥40	Velarino-Guell et al. 2014 Appel-Cresswell et al., 2014
DNAJC6 (PARK19)	Juvenile-onset Poor response to Levodopa Pyramidal signs Dystonia Seizures Mental retardation Fast progression	≤10	Edvardson et al. 2012 Köroğlu et al. 2013 Olgiati et al. 2016
GBA	Late-onset Typical Good response to Levodopa Cognitive and behavioural symptoms	≥60	Sidransky et al. 2009 (Sidransky and López, 2012)
PANK2	Early-onset Dystonia	≥20	Zhou et al. 2001 Schneider et al. 2012
PLA2G6 (PARK14)	Early-onset Atypical Fast progression	≥30	Paisan -Ruiz et al. 2009 Schneider et al. 2012
PODXL	Juvenile-onset Good response to Levodopa Dyskinesia Off-dystonia	≤20	(Sudhaman et al., 2016)
PTRHD1	Early-onset	≤30	Khodadadi et al., 2017

Gene (locus)	Typical symptoms	Age of onset	Reference
	Good response to Levodopa Pyramidal signs Dystonia Mental retardation Muscular weakness		
RAB39B	Early-onset PD Dopaminergic neuronal loss Lewy bodies Good response to Levodopa Mental retardation X-linked inheritance	≤40	Wilson et al. 2014 Lesage et al. 2015
SPG11	Early-onset Spastic paraplegia	≤50	Guidubaldi et al. 2011
TMEM230	Early-onset Autosomal dominant Lewy bodies present Good response to Levodopa Slow progression	≤50	HX Deng 2016 DA Olszewska 2016 Deng et al 2016 Mandemakers et al., 2017 Giri et al., 2017 Yan et al., 2017
UCHL1 (PARK5)	Early-onset Autosomal dominant Slow progression	≤50	Leroy et al. 1998 Maraganore, et al.1999 (Ragland et al., 2009)
VPS13C	Early-onset Dopamine neuronal loss Good response to Levodopa	≤50	Lesage et al 2016

Gene (locus)	Typical symptoms	Age of onset	Reference
	Rigidity Resting tremor Fast Progression		

NR: Not reported

Table 2 A summary of all published mutation screening studies performed on SA PD patients

Gene	Patients	Methods	Main findings	Reference
Autosomal Dominant Genes				
ATXN2 (Ataxin 2)	398	Polyglutamine repeat expansions (PCR, Capillary Gel Electrophoresis)	No pathogenic mutations were found	(Wang et al., 2015)
ATXN3 (Ataxin 3)	398	Polyglutamine repeat expansions (PCR, Capillary Electrophoresis)	No pathogenic mutations were found	(Wang et al., 2015)
CACNA1A (Calcium Voltage-Gated Channel Subunit Alpha1 A)	398	Polyglutamine repeat expansions (PCR, Capillary Electrophoresis)	No pathogenic mutations were found	(Wang et al., 2015)
EIF4G1 (Eukaryotic Translation Initiation Factor 4G1)	418	R1205H mutation (KASP genotyping assay)	p.R1205H mutation was not present	(Blanckenberg et al., 2014)

Gene	Patients	Methods	Main findings	Reference
GBA (Glucocerebrosidase Beta)	105	Exons 8, 9, 10 and11 were screened (PCR, Capillary Electrophoresis)	Severe mutations, like p.L444P were absent	(Barkhuizen et al., 2017)
GCH1 (GTP Cyclohydrolase 1)	88	Exons 1,2,3,5 and 6 (Exon dosage analysis using MLPA)	No pathogenic mutations were found	(Keyser et al., 2010)
LRRK2 (Leucine-Rich Repeat Kinase 2)	205	G2019S mutation (HRM analysis)	Four probands harboured the mutation in a heterozygous state	(Bardien et al., 2010)
	88	Exons 1,2,10,15,27,41,49 (Exon dosage analysis using MLPA)	No pathogenic mutations were found	(Keyser et al., 2009a)
	23	All 51 exons (sequencing)	Four rare non-synonymous variants unique to patients with PD in SA were observed p.I610T p.H1758P p.N2133S p. T2324S	(Trinh et al., 2015)
	210	Exons 1, 2, 8, 10, 15, 27,49 and G2019S.	No pathogenic mutations were found	(van der Merwe et al., 2016)

Gene	Patients	Methods	Main findings	Reference
		(Exon dosage analysis using MLPA))		
	16	G2019S. (MLPA and CNV)	No pathogenic mutations were found	(Mahne et al., 2016)
SNCA (Alpha-Synuclein)	88	All six exons and A30P (exon dosage analysis using MLPA)	One patient had a whole-gene triplication	(Keyser et al., 2009)
	210	All six exons and A30P (Exon dosage analysis using MLPA)	No pathogenic mutations were found	(van der Merwe et al., 2016)
	16 Black patients	All six exons and A30P (Exon dosage analysis using MLPA)	No pathogenic mutations were found	(Mahne et al., 2016)
TBP (TATA box binding protein)	398	Polyglutamine repeat expansions (PCR, Capillary Electrophoresis)	No pathogenic mutations were found	(Wang et al., 2015)
VPS35 (Vascular Protein Sorting-associated protein 35)	418	D620N Mutation (KASP genotyping assay)	p.D620N mutation was not present	(Blanckenberg et al., 2014)

Gene	Patients	Methods	Main findings	Reference
Autosomal Recessive Genes				
ATP13A2 (ATPase Type 13A2)	88	Exons 2 and 9 (dosage analysis MLPA)	No pathogenic mutations were found	(Keyser et al., 2009a)
DJ-1 (Daisuke-Junko-1/ *PARK7)*	30	All seven exons and 5'-UTR (dosage analysis MLPA)	16 bp deletion variant (g.-6_+10del) was identified and shown to reduce transcriptional activity by approximately 50% in cell culture systems	(Keyser et al., 2009b)
	150	5'UTR g.-6_+10del variant (ss831883147) (dosage analysis MLPA)	Two patients had the deletion	(Keyser et al., 2009b)
	402	5'UTR g.-6_+10del variant (ss831883147) (KASP genotyping assay)	One patient carried the deletion allele in a heterozygous state; the allele frequency was estimated to be 0,1 % (1/804). None of the 528 controls had this variant	(Glanzmann et al., 2013)
	402	5'UTR g.168_185del variant (ss831883149) (HRM)	Present in <1% of both cases and controls	(Glanzmann et al., 2013)
PARK2	91	All 12 exons	No pathogenic mutations were found	(Bardien et al., 2009)

Gene	Patients	Methods	Main findings	Reference
(Parkin)		(SSCP and HRM)		
	229	All 12 exons (HRM and MLPA)	Heterozygous, homozygous and four novel mutations in the Parkin gene were found. P113fsX163 deletion of exon 3 G430D deletion of exon 4 Duplication of exons 2e6 duplication of exon 5 Silent substitutions (p.P37P, p.C238C, p.L261L, p.A397A and p.R402R) Polymorphisms (p.A82E and p.P437L) Variants found in matched controls at MAF ≥1% (p.Q34R, p.S167N, p.M192L, p.R334C, p.V380L, p.D394N).	(Haylett et al., 2012)
	88	All 12 exons (dosage analysis MLPA)	Heterozygous duplication of exon 2 Heterozygous deletion of exon 9	(Keyser et al., 2009a)

Gene	Patients	Methods	Main findings	Reference
			Heletion of exon 4 in another proband. Heterozygous p.G430D mutation	
	3	R275W, M432V (WES and MLPA)	Identified both homozygous and heterozygous forms of p.R275W, p.M432V variants	(Carr et al., 2016)
	210	All 12 exons (MLPA and CNV)	Exon 4 deletion p.M192L polymorphism found in 13 individuals	(van der Merwe et al., 2016)
	16	All 12 exons (MLPA and CNV)	One heterozygous deletion of exon 4	(Mahne et al., 2016)
PINK1 (PTEN-induced putative kinase 1)	154	All eight exons (HRM and sequencing)	16 sequence variants found 13 were polymorphisms (of which five were novel) One homozygous mutation p.Y258X One novel heterozygous variant p.P305A One heterozygous variant p.E476K)	(Keyser et al., 2010a)
	88	All eight exons (dosage analysis MLPA)	No pathogenic mutations were found	(Keyser et al., 2009a)

Gene	Patients	Methods	Main findings	Reference
	210	All eight exons (MLPA and CNV)	No pathogenic mutations were found	(van der Merwe et al., 2016)

AD, Autosomal Dominant; AR, Autosomal Recessive; CNV, Copy Number Variation; HRM, High Resolution Melt; MLPA, Multiplex Ligation-dependent Probe Amplification; SSCP, Single Strand Conformation Polymorphism Analysis; KASP, Kompetitive Allele Specific Polymerase Chain Reaction; UTR, Untranslated Region; WES, Whole Exome Sequencing

Table 3 HRM control screening results

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
67,86	Caucasian	Female	46	Absent	Absent	Absent	**Absent**	Absent
68,68	Caucasian	Female	44	Absent	Absent	Absent	Absent	Absent
68,71	Caucasian	Male	68	Absent	Absent	Absent	Absent	Absent
68,92	Caucasian	Female	81	Absent	Absent	Absent	Absent	Absent
68,93	Caucasian	Female	68	Absent	Absent	Absent	Absent	Absent
68,94	Caucasian	Male	63	Absent	Absent	Absent	Absent	Absent
68,95	Caucasian	Male	70	Absent	Absent	Absent	Absent	Absent
69,16	Caucasian	Male	52	Absent	Absent	Absent	Absent	**Absent**
69,17	Caucasian	Female	53	Absent	**Absent**	Absent	Absent	Absent
69,38	Caucasian	Male	78	Absent	Absent	Absent	Absent	Absent
69,39	Caucasian	Male	51	Absent	Absent	**Absent**	Absent	Absent
69,40	Caucasian	Male	64	Absent	**Absent**	Absent	**Absent**	Absent
69,41	Caucasian	Female	58	Absent	Absent	Absent	Absent	Absent
69,65	Caucasian	Male	67	Absent	Absent	Absent	Absent	Absent
69,66	Caucasian	Female	58	Absent	Absent	Absent	Absent	Absent
69,67	Caucasian	Male	71	Absent	Absent	Absent	Absent	Absent
69,81	Caucasian	Female	46	Absent	Absent	Absent	Absent	Absent
69,82	Caucasian	Male	74	Absent	Absent	Absent	Absent	Absent
69,84	Caucasian	Male	57	Absent	Absent	Absent	Absent	Absent
69,85	Caucasian	Male	46	Absent	Absent	**Absent**	Absent	Absent
69,86	Caucasian	Male	60	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
70,28	Caucasian	Female	62	Absent	Absent	Absent	Absent	Absent
70,44	Caucasian	Male	51	Absent	Absent	Absent	Absent	Absent
70,45	Caucasian	Male	72	Absent	Absent	Absent	Absent	Absent
70,58	Caucasian	Male	82	Absent	Absent	Absent	Absent	Absent
70,63	Caucasian	Female	39	Absent	Absent	**Absent**	Absent	Absent
70,64	Caucasian	Male	46	Absent	Absent	Absent	Absent	Absent
70,65	Caucasian	Female	47	Absent	Absent	**Absent**	Absent	**Absent**
70,97	Caucasian	Male	69	Absent	Absent	Absent	Absent	Absent
70,99	Caucasian	Male	60	Absent	Absent	Absent	Absent	Absent
71,41	Caucasian	Female	34	**Absent**	Absent	Absent	Absent	Absent
71,42	Caucasian	Female	35	**Absent**	Absent	Absent	Absent	Absent
71,43	Caucasian	Male	61	**Absent**	**Absent**	Absent	**Absent**	**Absent**
71,59	Caucasian	Female	56	Absent	Absent	Absent	**Absent**	Absent
71,60	Caucasian	Female	49	Absent	Absent	Absent	Absent	Absent
71,79	Caucasian	Male	64	Absent	Absent	Absent	Absent	Absent
71,80	Caucasian	Male	56	Absent	Absent	Absent	Absent	Absent
71,82	Caucasian	Female	64	Absent	Absent	Absent	Absent	Absent
72,43	Caucasian	Female	70	Absent	Absent	Absent	Absent	Absent
72,45	Caucasian	Male	62	Absent	Absent	**Absent**	Absent	Absent
72,52	Caucasian	Female	40	Absent	**Absent**	Absent	**Absent**	Absent
72,53	Caucasian	Male	37	Absent	Absent	Absent	Absent	**Absent**
72,55	Caucasian	Male	50	**Absent**	Absent	Absent	**Absent**	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
72,57	Caucasian	Male	41	Absent	Absent	Absent	Absent	Absent
72,58	Caucasian	Male	38	**Absent**	**Absent**	Absent	**Absent**	Absent
72,61	Caucasian	Male	69	Absent	Absent	Absent	Absent	Absent
72,98	Caucasian	Male	75	Absent	Absent	Absent	Absent	Absent
73,00	Caucasian	Female	33	Absent	Absent	Absent	Absent	Absent
73,05	Caucasian	Male	61	Absent	Absent	Absent	Absent	Absent
73,08	Caucasian	Male	63	**Absent**	Absent	Absent	Absent	Absent
73,13	Caucasian	Female	50	Absent	Absent	Absent	Absent	Absent
73,15	Caucasian	Female	74	Absent	Absent	Absent	Absent	Absent
73,16	Caucasian	Female	57	Absent	Absent	Absent	Absent	Absent
73,18	Caucasian	Male	67	Absent	Absent	Absent	Absent	Absent
73,33	Caucasian	Male	70	Absent	Absent	Absent	Absent	Absent
73,35	Caucasian	Female	63	Absent	Absent	Absent	Absent	Absent
73,36	Caucasian	Female	53	Absent	Absent	Absent	Absent	Absent
73,38	Caucasian	Female	74	Absent	Absent	Absent	Absent	Absent
73,42	Caucasian	Female	48	Absent	Absent	Absent	Absent	**Absent**
73,44	Caucasian	Male	51	Absent	Absent	Absent	Absent	**Absent**
73,45	Caucasian	Female	69	Absent	Absent	Absent	Absent	Absent
74,14	Caucasian	Male	40	Absent	Absent	Absent	Absent	Absent
74,16	Caucasian	Male	58	Absent	Absent	Absent	Absent	Absent
74,37	Caucasian	Female	38	**Absent**	Absent	Absent	Absent	Absent
75,68	Caucasian	Female	41	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
75,69	Caucasian	Male	50	Absent	Absent	**Absent**	Absent	Absent
75,71	Caucasian	Male	58	Absent	Absent	Absent	Absent	Absent
75,74	Caucasian	Male	64	Absent	Absent	Absent	**Absent**	Absent
75,76	Caucasian	Female	49	Absent	Absent	Absent	Absent	Absent
75,78	Caucasian	Male	79	Absent	Absent	Absent	Absent	Absent
75,79	Caucasian	Male	64	Absent	Absent	Absent	**Absent**	Absent
75,82	Caucasian	Male	76	Absent	**Absent**	Absent	Absent	**Absent**
75,85	Caucasian	Male	46	Absent	Absent	Absent	Absent	**Absent**
76,22	Caucasian	Male	41	Absent	Absent	Absent	Absent	Absent
76,24	Caucasian	Male	66	Absent	Absent	Absent	Absent	Absent
76,25	Caucasian	Male	70	Absent	Absent	Absent	Absent	Absent
77,10	Caucasian	Female	50	Absent	Absent	**Absent**	Absent	Absent
78,33	Caucasian	Male	63	Absent	Absent	Absent	Absent	Absent
78,35	Caucasian	Male	41	Absent	Absent	Absent	Absent	**Absent**
78,42	Caucasian	Female	29	Absent	Absent	Absent	Absent	Absent
86,90	Caucasian	Male	47	Absent	Absent	Absent	Absent	Absent
86,97	Caucasian	Male	54	Absent	Absent	Absent	Absent	Absent
86,98	Caucasian	Female	49	Absent	Absent	Absent	Absent	Absent
87,01	Caucasian	Female	43	Absent	Absent	Absent	Absent	Absent
87,04	Caucasian	Male	35	Absent	Absent	Absent	Absent	Absent
87,06	Caucasian	Male	59	Absent	Absent	Absent	Absent	Absent
87,08	Caucasian	Female	44	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
87,09	Caucasian	Female	62	Absent	Absent	**Absent**	Absent	Absent
87,12	Caucasian	Male	61	Absent	Absent	Absent	Absent	Absent
87,16	Caucasian	Female	47	Absent	Absent	Absent	**Absent**	Absent
87,17	Caucasian	Male	59	Absent	**Absent**	Absent	Absent	Absent
87,18	Caucasian	Male	72	Absent	Absent	Absent	Absent	Absent
87,20	Caucasian	Female	71	Absent	**Absent**	**Absent**	Absent	Absent
87,21	Caucasian	Male	46	**Absent**	**Absent**	Absent	Absent	Absent
87,22	Caucasian	Male	48	**Absent**	**Absent**	Absent	Absent	**Absent**
87,23	Caucasian	Female	29	**Absent**	Absent	**Absent**	Absent	Absent
50,87	Afrikaner	Male	87	Absent	Absent	Absent	Absent	**Absent**
51,68	Afrikaner	Female	104	Absent	Absent	Absent	Absent	Absent
52,25	Afrikaner	Female	101	Absent	Absent	Absent	Absent	Absent
52,37	Afrikaner	Female	84	Absent	Absent	Absent	Absent	Absent
52,96	Afrikaner	Male	48	Absent	Absent	Absent	Absent	Absent
52,97	Afrikaner	Female	89	Absent	Absent	Absent	Absent	**Absent**
52,98	Afrikaner	Male	85	**Absent**	**Absent**	Absent	Absent	Absent
52,99	Afrikaner	Female	105	Absent	Absent	Absent	**Absent**	Absent
53,00	Afrikaner	Female	93	Absent	Absent	Absent	Absent	Absent
53,01	Afrikaner	Male	92	Absent	Absent	Absent	Absent	**Absent**
53,02	Afrikaner	Female	97	**Absent**	Absent	Absent	Absent	Absent
53,03	Afrikaner	Female	92	Absent	Absent	Absent	Absent	Absent
53,04	Afrikaner	Female	86	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
53,14	Afrikaner	Female	93	Absent	Absent	Absent	Absent	Absent
53,15	Afrikaner	Female	87	Absent	Absent	Absent	Absent	Absent
53,16	Afrikaner	Female	91	Absent	Absent	Absent	Absent	Absent
53,43	Afrikaner	Male	58	Absent	Absent	Absent	**Absent**	Absent
53,68	Afrikaner	Female	92	Absent	Absent	Absent	**Absent**	Absent
53,69	Afrikaner	Female	89	Absent	Absent	**Absent**	Absent	Absent
53,70	Afrikaner	Female	92	Absent	Absent	Absent	Absent	Absent
54,08	Afrikaner	Male	64	Absent	**Absent**	Absent	Absent	Absent
54,09	Afrikaner	Male	92	Absent	Absent	Absent	Absent	Absent
54,10	Afrikaner	Female	97	Absent	Absent	Absent	Absent	Absent
54,11	Afrikaner	Female	86	Absent	Absent	Absent	**Absent**	Absent
54,12	Afrikaner	Female	90	Absent	Absent	Absent	Absent	Absent
54,13	Afrikaner	Female	89	Absent	Absent	Absent	Absent	**Absent**
54,23	Afrikaner	Female	86	**Absent**	**Absent**	Absent	Absent	Absent
54,24	Afrikaner	Female	91	Absent	Absent	Absent	Absent	Absent
54,25	Afrikaner	Female	87	Absent	Absent	**Absent**	Absent	Absent
54,26	Afrikaner	Female	100	Absent	Absent	Absent	Absent	Absent
54,27	Afrikaner	Female	104	Absent	Absent	**Absent**	Absent	Absent
54,58	Afrikaner	Female	94	Absent	Absent	Absent	Absent	Absent
54,59	Afrikaner	Female	95	Absent	Absent	Absent	Absent	Absent
54,60	Afrikaner	Female	92	Absent	Absent	Absent	Absent	Absent
54,61	Afrikaner	Male	88	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
54,64	Afrikaner	Female	86	Absent	Absent	Absent	Absent	Absent
54,65	Afrikaner	Male	94	**Absent**	Absent	Absent	Absent	Absent
54,66	Afrikaner	Female	98	Absent	Absent	Absent	Absent	Absent
54,67	Afrikaner	Female	84	**Absent**	Absent	Absent	Absent	Absent
54,68	Afrikaner	Male	79	Absent	Absent	Absent	Absent	Absent
54,69	Afrikaner	Female	103	Absent	Absent	Absent	Absent	Absent
54,70	Afrikaner	Female	97	Absent	Absent	Absent	Absent	Absent
54,71	Afrikaner	Female	87	Absent	Absent	Absent	Absent	Absent
54,72	Afrikaner	Female	101	**Absent**	Absent	**Absent**	Absent	Absent
54,80	Afrikaner	Female	89	Absent	Absent	Absent	Absent	Absent
54,81	Afrikaner	Female	91	Absent	**Absent**	Absent	Absent	Absent
54,82	Afrikaner	Female	100	Absent	Absent	Absent	Absent	Absent
54,83	Afrikaner	Female	95	Absent	Absent	Absent	Absent	Absent
54,84	Afrikaner	Female	89	Absent	Absent	Absent	Absent	Absent
54,85	Afrikaner	Female	96	Absent	Absent	Absent	Absent	Absent
54,87	Afrikaner	Female	95	Absent	Absent	Absent	Absent	Absent
67,89	Afrikaner	Male	61	Absent	Absent	Absent	**Absent**	Absent
68,69	Afrikaner	Male	61	Absent	**Absent**	Absent	Absent	Absent
68,70	Afrikaner	Male	54	**Absent**	Absent	**Absent**	Absent	Absent
69,18	Afrikaner	Female	47	Absent	Absent	Absent	Absent	Absent
69,19	Afrikaner	Male	83	Absent	Absent	Absent	Absent	Absent
69,20	Afrikaner	Male	55	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
69,45	Afrikaner	Male	38	Absent	Absent	Absent	**Absent**	Absent
69,46	Afrikaner	Male	35	**Absent**	Absent	**Absent**	Absent	Absent
69,47	Afrikaner	Female	73	Absent	Absent	Absent	Absent	Absent
70,60	Afrikaner	Female	47	Absent	Absent	Absent	Absent	Absent
70,98	Afrikaner	Female	47	Absent	**Absent**	Absent	**Absent**	Absent
71,00	Afrikaner	Male	55	Absent	Absent	Absent	Absent	Absent
71,37	Afrikaner	Female	58	Absent	Absent	Absent	Absent	**Absent**
71,38	Afrikaner	Male	72	Absent	Absent	Absent	Absent	Absent
71,39	Afrikaner	Female	41	**Absent**	Absent	**Absent**	Absent	Absent
71,40	Afrikaner	Male	52	Absent	Absent	Absent	Absent	**Absent**
71,48	Afrikaner	Female	37	Absent	Absent	Absent	Absent	Absent
71,83	Afrikaner	Female	43	Absent	**Absent**	Absent	Absent	Absent
72,46	Afrikaner	Male	69	Absent	Absent	Absent	Absent	Absent
72,50	Afrikaner	Male	56	Absent	Absent	Absent	Absent	Absent
73,01	Afrikaner	Male	61	Absent	Absent	Absent	Absent	Absent
73,12	Afrikaner	Male	65	Absent	Absent	Absent	Absent	Absent
73,32	Afrikaner	Male	57	Absent	**Absent**	Absent	Absent	Absent
73,34	Afrikaner	Male	39	Absent	Absent	Absent	Absent	Absent
73,41	Afrikaner	Male	43	Absent	Absent	Absent	Absent	**Absent**
73,46	Afrikaner	Male	55	Absent	Absent	Absent	Absent	Absent
74,15	Afrikaner	Male	43	Absent	Absent	Absent	Absent	**Absent**
74,20	Afrikaner	Female	73	Absent	Absent	Absent	Absent	Absent

Control Information				**HRM Screening Result**				
Lab ID	**Ethnicity**	**Gender**	**Age in 2018**	**Gene**				
				CFAP65 c.A3151G	***RFT1*** c.C1450G	***NRXN2*** c.G3008A	***TEP1*** c.A1276G	***CCNF*** c.G1176C
74,23	Afrikaner	Male	64	Absent	Absent	Absent	Absent	Absent
74,38	Afrikaner	Female	51	Absent	Absent	**Absent**	Absent	Absent
75,77	Afrikaner	Male	38	Absent	Absent	Absent	Absent	Absent
75,81	Afrikaner	Male	55	Absent	Absent	Absent	Absent	Absent
76,21	Afrikaner	Male	61	Absent	Absent	Absent	Absent	Absent
76,26	Afrikaner	Male	68	Absent	Absent	Absent	Absent	Absent
77,12	Afrikaner	Male	61	Absent	Absent	**Absent**	Absent	Absent
77,39	Afrikaner	Female	73	Absent	Absent	Absent	Absent	Absent
77,40	Afrikaner	Male	59	Absent	Absent	Absent	Absent	Absent
78,30	Afrikaner	Male	28	**Absent**	Absent	Absent	Absent	**Absent**
78,34	Afrikaner	Male	52	Absent	Absent	**Absent**	Absent	Absent
78,40	Afrikaner	Male	34	Absent	Absent	Absent	Absent	Absent
86,89	Afrikaner	Male	68	Absent	Absent	Absent	Absent	**Absent**
86,92	Afrikaner	Male	54	Absent	**Absent**	Absent	Absent	Absent
86,95	Afrikaner	Male	28	**Absent**	Absent	Absent	**Absent**	Absent
86,99	Afrikaner	Female	30	**Absent**	Absent	Absent	**Absent**	Absent
87,03	Afrikaner	Female	43	**Absent**	**Absent**	Absent	**Absent**	Absent

Absent, Sanger sequence verified HRM result.

2. Appendix II

2. Supplementary Figures

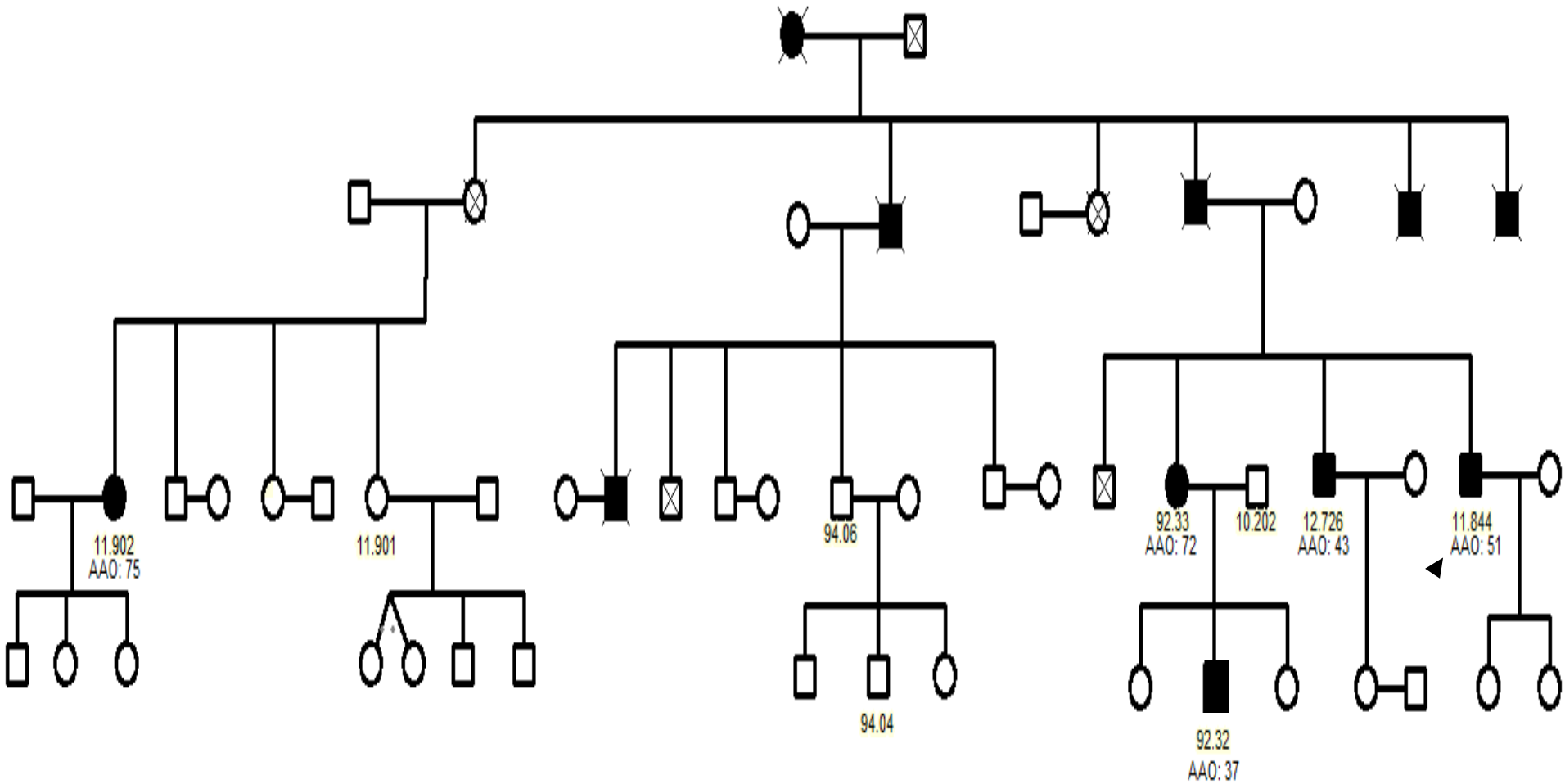


Figure 1 Full pedigree of family ZA253. Circles denote females and squares depict males. The filled in symbols indicate affected individuals. The intersecting diagonal lines indicate that the person is deceased. The numbers below each individual is the laboratory ID number. For readability and confidentiality, the pedigree is greatly simplified. Branches without medically confirmed PD or without DNA samples were omitted. AAO, Age at onset, the black arrow shows the proband.

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Ethics Letter

16-Mar-2018

Ethics Reference #: 2002/C059

Title: Genetic analysis of inherited Parkinson's Disease and other related movement disorders

Dear Prof Jonathan Carr,

Your request for extension/annual renewal of ethics approval dated 19 February 2018 refers.

The Health Research Ethics Committee reviewed and approved the annual progress report you submitted through an expedited review process.

The approval of the research project is extended for a further year.

Approval Date: 16 March 2018

Expiry Date: 15 March 2019

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly submit **ONE HARD COPY** to Elvira Rohland, RDSD, Room 5007, Teaching Building, and **ONE ELECTRONIC COPY** to ethics@sun.ac.za .

Please remember to use your **protocol number (2002/C059)** on any documents or correspondence with the HREC concerning your research protocol.

National Health Research Ethics Council (NHREC) Registration Numbers: REC-130408-012 for HREC1 and REC-230208-010 for HREC2

Federal Wide Assurance Number: 00001372

Institutional Review Board (IRB) Number: IRB0005240 for HREC1
Institutional Review Board (IRB) Number: IRB0005239 for HREC2

Fakulteit Geneeskunde en Gesondheidswetenskappe
Faculty of Medicine and Health Sciences

Afdeling Navorsingsontwikkeling en -Steun • Research Development and Support Division

Posbus/PO Box 241 • Cape Town 8000 • Suid-Afrika/South Africa
Tel: +27 (0) 21 938 9677

Figure 2 Ethics certificate. The researcher was granted ethical approval and permission to conduct this study by the Health Research Ethics Committee (Stellenbosch University).